GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 3, 2003, 21:11:56 ; Search time 343.009 Seconds
(without alignments)
16426.693 Million cell updates/sec

Title: US-10-027-000-3
Perfect score: 2502
Sequence: 1 atggctgatattgatgttga.....attggtggtctggcgtgtaa 2502

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	260.8	10.4	588	21	AAF15000	Trichoderma reesei
2	184.4	7.4	3241	24	AAS18442	Contig 93 DNA enco
3	179	7.2	2401	21	AAA75634	Nucleotide sequenc
4	179	7.2	2401	21	AAZ56002	Contig 001 from co
5	177.4	7.1	1145	16	AAT04785	Cellobiase gene fr
6	175.2	7.0	2430	21	AAZ87294	S. venezuelae macr
7	175.2	7.0	2430	24	AAD39052	Streptomyces venez
8	175.2	7.0	12441	21	AAZ87284	S. venezuelae deso
9	175.2	7.0	13613	24	AAD39043	Streptomyces venez
10	167.2	6.7	13613	21	AAZ87319	S. venezuelae deso
11	161	6.4	2256	17	AAT32999	Chimaeric thermost
c 12	138.6	5.5	50937	21	AAA09469	Streptococcus olea
13	135.8	5.4	2166	18	AAT93682	Thermotoga maritim
14	135.8	5.4	2166	19	AAV36911	Thermotoga maritim
15	111.6	4.5	4403765	22	AAI99683	Mycobacterium tube
16	111.6	4.5	4411529	22	AAI99682	Mycobacterium tube
c 17	103	4.1	985	24	ABQ44816	Oligonucleotide fo
18	103	4.1	985	24	ABQ44817	Oligonucleotide fo
19	101.4	4.1	16836	19	AAV52831	Acetobacter xylinu
20	98.6	3.9	3849	22	AAF25795	S. chrysomallus ac
21	93.4	3.7	2271	11	AAQ06844	Amylase gene from
22	91.2	3.6	2291	9	AAN80309	Entire amylase gen
23	85.8	3.4	12588	15	AAQ63293	Sequence encoding
24	85.4	3.4	65140	22	AAD17184	Streptomyces nours
25	85.4	3.4	125401	22	AAD17186	Streptomyces nours
26	84.4	3.4	390	13	AAQ21833	Randomising oligon
27	84.4	3.4	390	14	AAQ36859	PCR primer for 5'
28	84.4	3.4	390	22	AAF76910	Sequence containin
29	84.4	3.4	390	24	AAI72775	Oligo #7 for cloni
30	83.8	3.3	3032	13	AAQ25325	Trichoderma reesei
31	83.8	3.3	3033	21	AAA63953	DNA encoding a bet
32	83.8	3.3	3033	21	AAZ38879	Trichoderma reesei
33	80.4	3.2	1140	19	AAV41733	Codon-optimised RA
34	80.4	3.2	7498	24	ABK91527	Modified HIV prote
35	80.2	3.2	1661	22	AAH21757	Corynebacterium gl
36	78.6	3.1	2064	14	AAQ52638	Streptomyces fradi
37	78.6	3.1	3269	16	AAT06024	cDNA encoding aven
38	78.2	3.1	1294	20	AAZ06824	Streptomyces albid
39	78.2	3.1	1294	22	AAH74537	Nucleotide sequenc
40	78.2	3.1	1521	16	AAQ85043	3-Hydroxysteroid-o
41	78.2	3.1	1521	17	AAT62676	3-hydroxysteroid o
42	78.2	3.1	1521	19	AAV11824	Streptomyces sp. s
43	78.2	3.1	1647	16	AAQ85042	3-Hydroxysteroid-o
44	78.2	3.1	1647	17	AAT62675	3-hydroxysteroid o
45	78.2	3.1	1647	19	AAV11823	Streptomyces sp. s

ALIGNMENTS

RESULT 1
AAF15000
ID AAF15000 standard; cDNA; 588 BP.
XX
AC AAF15000;
XX
DT 13-MAR-2001 (first entry)
XX
DE Trichoderma reesei EST SEQ ID NO:7523.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Trichoderma reesei.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.

XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 89; Page 3038-3039; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 588 BP; 130 A; 167 C; 157 G; 109 T; 25 other;

Query Match 10.4%; Score 260.8; DB 21; Length 588;
Best Local Similarity 92.8%; Pred. No. 4.7e-39;
Matches 283; Conservative 0; Mismatches 20; Indels 2; Gaps 1;

```
Qy   1870 ATCCAGGCCTGGTACGGCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGC 1929
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      1 ATCCAGGCCTGGTACGGCGGNAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGC 60

Qy   1930 GACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACAACCCC 1989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACAACCCC 120

Qy   1990 GCGTTTCTCAACTTCCGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTCTACGTC 2049
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCGTTTCTCAACTTCCGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTCTACGTC 180

Qy   2050 GGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTG 2109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTG 240

Qy   2110 TCCTACACCACTTTT--GCCTTTTCCAATCTCTCCGTGTCTCACAAGGACGGCAAGCTGA 2167
          ||||||||||||||   | |||||||||||||  |  |  ||  ||  | ||   |  |
Db    241 TCCTACACCACTTTTTGCCTTTTTCCAATCTCTTCCGGGGTCTTAACAANGGACGGNAAA 300

Qy   2168 GCGTG 2172
          || ||
Db    301 GCTTG 305
```

RESULT 2
AAS18442
ID AAS18442 standard; DNA; 3241 BP.
XX
AC AAS18442;
XX
DT 12-MAR-2002 (first entry)
XX
DE Contig 93 DNA encoding S. narbonensis polyketide synthase.
XX
KW Narbonolide polyketide synthase; PKS; narbomycin modification enzyme;
KW erythromycin; rapamycin; tylosin; picromycin; methylmycin; neomethymycin;
KW agriculture; ds.
XX
OS Streptomyces narbonensis.
XX
PN US6303767-B1.
XX
PD 16-OCT-2001.
XX
PF 05-NOV-1999; 99US-0434288.
XX
PR 05-NOV-1998; 98US-107093P.
PR 27-MAY-1999; 99US-0320878.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Betlach MC, McDaniel R;
XX
DR WPI; 2002-065495/09.
XX
PT Nucleic acids encoding narbonolide polyketide synthases from
PT Streptomyces narbonensis, useful for the recombinant production of
PT polyketides, e.g. narbomycin -
XX
PS Claim 1; Column 20-22; 24pp; English.
XX
CC The present invention relates to recombinant DNA vectors (cosmids)
CC that encode for the narbonolide polyketide synthase (PKS) enzyme and
CC various narbomycin modification enzymes from Streptomyces narbonensis.
CC The recombinant DNA vectors can be used to produce recombinant ketide
CC synthases and a variety of different polyketides (e.g. erythromycin,
CC rapamycin, tylosin, narbomycin, picromycin, methylmycin and
CC neomethymycin) for use in agriculture, medicine and health. The
CC recombinant vectors may be used to produce polyketides in relatively
CC high yields. AAS18432-AAS18443 represent contig DNA sequences that
CC encode for S. narbonensis PKS enzymes.
XX
SQ Sequence 3241 BP; 520 A; 1174 C; 1125 G; 422 T; 0 other;

Query Match 7.4%; Score 184.4; DB 24; Length 3241;
Best Local Similarity 48.2%; Pred. No. 5.8e-25;
Matches 1028; Conservative 0; Mismatches 956; Indels 150; Gaps 12;

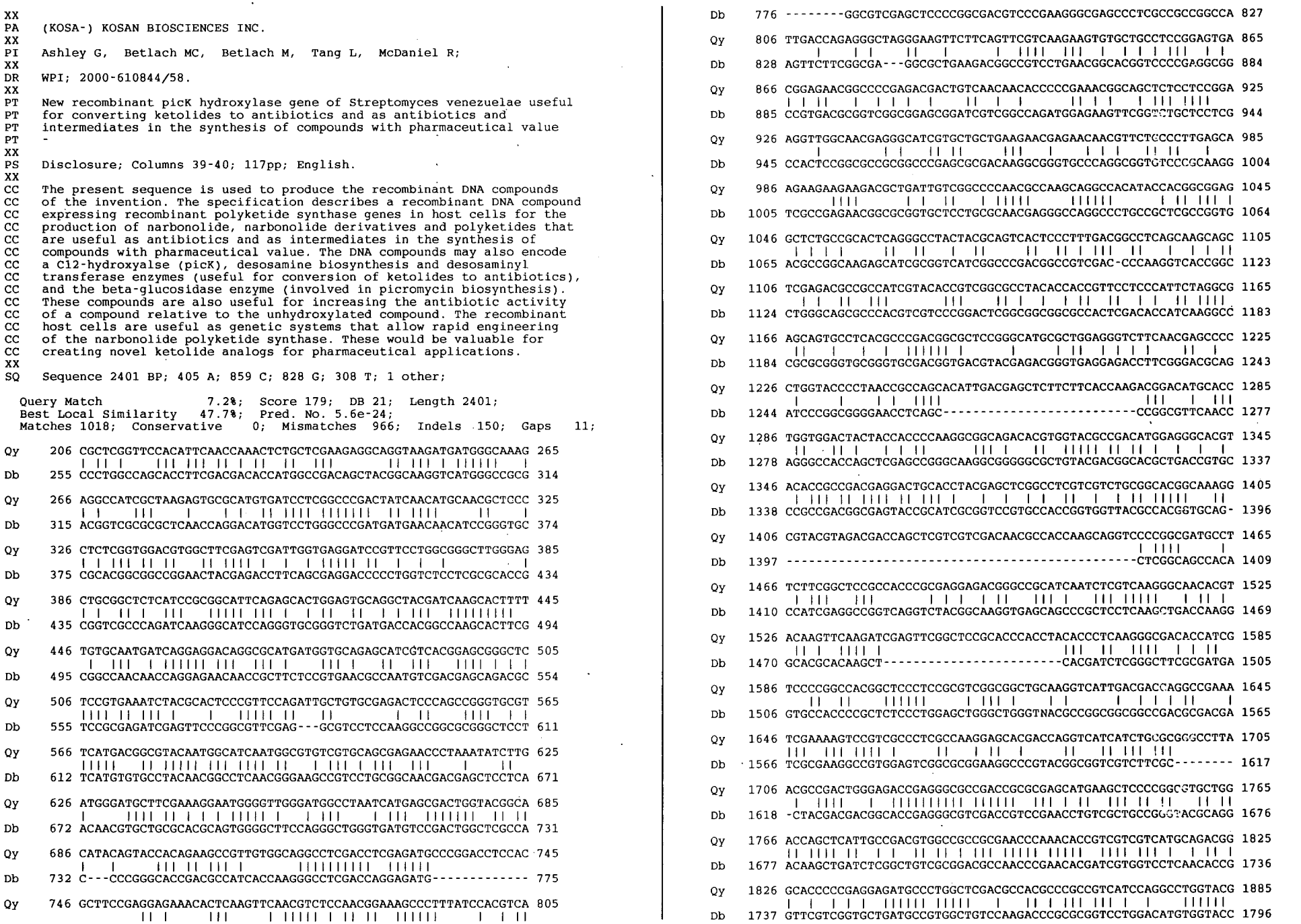
```
Qy    206 CGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAG 265
          | || |    ||| ||| || |  |  ||  |||       || | | | ||||||   |
Db   1151 CCCTGGCCAGCACCTTCGACGACTCCATGGCCGACAGCTACGGCAGGGTCATGGGCCGCG 1210

Qy    266 AGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCC 325
          | |    |||   | |   | |  || | || |||||||  || ||||     ||    |
Db   1211 ACGGACGCGCGCTGGGCCAGGACATGGTTCTGGGCCCGATGATGAACAACATCCGGGTGC 1270

Qy    326 CTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAG 385
          | | ||| || ||   || |||| |  |  | ||||| ||  |  | |.         |
Db   1271 CACACGGCGGCCGGAACTACGAGACCTTCAGCGAGGACCCCCTGGTCTCCTCGCGCACCG 1330

Qy    386 CTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTT 445
          | |  || |  |||   ||||| ||| |  | ||  ||  | | |||  |||||||||
Db   1331 CGGTCGCCCAGATCAAGGGCATCCAGGGTGCGGGTCTGATGACCACGGCCAAGCACTTCG 1390

Qy    446 TGTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGCATCGTCACGGAGCGGGCTC 505
          |  |||  | ||||||  |||  |||  | |  ||  |   ||  |||   |||| | | |
Db   1391 CGGCCAACAACCAGGAGAACAACCGCTTCAGCGTCAACGCCACGGTCGACGAGCAGACGC 1450

Qy    506 TCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCCAGCCGGGTGCGT 565
          |||| || ||| |    |   ||||| ||  ||         |  ||    |||   | |
Db   1451 TCCGCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCGAAGGCCGGCGCGGCCTCCT 1507

Qy    566 TCATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGCGAGAACCCTAAATATCTTG 625
          |||||   || || || ||| |||| |||  | ||| | |||  |||     |    ||
Db   1508 TCATGTGTGCCTATAACGGCGTCAACGGCAAGCCGTCCTGCGGCAACGACGAGCTGCTCA 1567

Qy    626 ATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCGACTGGTACGGCA 685
```

Cannot use this due to In re Durand

```
Db   1568 ACAACGTGCTGCGCACGCAGTGGGGCTTCCAGGGCTGGGTGATGTCCGACTGGCTCGCCA 1627

Qy    686 CATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGCCCGGACCTCCAC 745

Db   1628 C---CCCGGGCACGGACGCCATCACCAAGGGCCTCGACCAGGAGAT----GGGCGTCGAG 1680

Qy    746 GCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCTTTATCCACGTCA 805

Db   1681 CTCCCCGGCGACATCCCGCCGGGCGAGCCCTCGCCGCCGGCCAAGTTCTTCGGTGACGCG 1740

Qy    806 TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTGCTGCCTCCGGAGTGA 865

Db   1741 CTGA--------------------AGCAGGCCGTCCTGAACGGCACGGTCCCCGAGGCGG 1780

Qy    866 CGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCCGAAACGGCAGCTCTCCTCCGGA 925

Db   1781 CCGTGACGCGGTCGGCGGAGCGCATCGTCAACCAGATGGACAAGTTCGGTCTGCTCCTCG 1840

Qy    926 AGGTTGGCAACGAGGGCATCGTGCTGCTGAAGAACGAGAACAACGTTCTGCCCTTGAGCA 985

Db   1841 CGACTCCGGCGCCCCGCCCCGAGCGTGACAAGGCGGGCGCCCAGGCGGTGTCCCGCAAGG 1900

Qy    986 AGAAGAAGAAGACGCTGATTGTCGGCCCCAACGCCAAGCAGGCCACATACCACGGCGGAG 1045

Db   1901 TCGCCGAGAACGGCGCGGTGCTCCTGCGCAACGAGGGCCAGGCCCTGCCGCTCGCCGGTG 1960

Qy   1046 GCTCTGCCGCACTCAGGGCCTACTACGCAGTCACTCCCTTTGACGGCCTCAGCAAGCAGC 1105

Db   1961 ACGCCGGCAAGAGCATCGCCGTCATCGGCCCGACGGCCGTCGAC-CCCAAGGTCACCGGC 2019

Qy   1106 TCGAGACGCCGCCATCGTACACCGTCGGCGCCTACACCACCGTTCCTCCCATTCTAGGCG 1165

Db   2020 CTGGGCAGCGCCCACGTCGTCCCGGACTCGGCGGCGGCGCCGCTCGACACCATCAAGGCC 2079

Qy   1166 AGCAGTGCCTCACGCCCGACGGCGCTCCGGGCATGCGCTGGAGGGTCTTCAACGAGCCCC 1225

Db   2080 CGCGCGGGCGCGGGTGCGACGGTGACGTACGAGACGGGTGAGGAGACCTTCGGGACGCGG 2139

Qy   1226 CTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCTTCACCAAGACGGACATGCACC 1285

Db   2140 ATCCCGGCGGCGCAGCTCAGC------------------------CCGGCGTTCAACC 2173

Qy   1286 TGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGGTACGCCGACATGGAGGGCACGT 1345

Db   2174 AGGGCCACCAGCTGGAGCCGGGCAAGGCGGGGGCGCTGTACGACGGCACGCTGACCGTGC 2233

Qy   1346 ACACCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGTCGTCTGCGGCACGGCAAAGG 1405

Db   2234 CCGCCGACGGCGAGTACCGCATCTCGGTCAAGGCCACCGGTGGCTACGCGACGGTGCAG- 2292

Qy   1406 CGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCCCCGGCGATGCCT 1465

Db   2293 ---------------------------------------------CTCGGCAGCCACA 2305

Qy   1466 TCTTCGGCTCCGCCACCCGCGAGGAGACGGGCCGCATCAATCTCGTCAAGGGCAACACGT 1525

Db   2306 CCATCGAGGCCGGTCAGGTCTACGGCAAGGTGAGCAGCCCGCTCCTCAAGCTGACCAAGG 2365

Qy   1526 ACAAGTTCAAGATCGAGTTCGGCTCCGCACCCACCTACACCCTCAAGGGCGACACCATCG 1585

Db   2366 GCACGCACAAGCT----------------------CACGATCTCGGGCTTCGCGATGA 2401

Qy   1586 TCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAAA 1645

Db   2402 GCGCCACGCCGCTCTCCCTGGAGCTGGGCTGGGTGACGCCGGAGGCAGCCGACGCGACGA 2461

Qy   1646 TCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTTA 1705

Db   2462 TCGCGAAGGCCGTGGAGTCGGCGCGGAAGGCCCGTACGGCCATCGTGTTCGCG------- 2514

Qy   1706 ACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGG 1765

Db   2515 --TACGACGACGGCACCGAGGGCGTCGACCGTCCGAACCTGTCGCTGCCGGGTACGCAGG 2572

Qy   1766 ACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGG 1825

Db   2573 ACAAGCTGATCTCGGCGGTCGCCGACGCGAACCCGAACACGATCGTGGTCCTCAACACCG 2632

Qy   1826 GCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACG 1885

Db   2633 GTTCGTCGGTGCTGATGCCGTGGCTGTCCAAGACCCGCGCGGTCCTGGACATGTGGTACC 2692

Qy   1886 GCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGG 1945

Db   2693 CGGGCCAGGCGGGCGCCGAGGCGACCGCCGCGCTGCTCTACGGTGACGTGAACCCGAGCG 2752

Qy   1946 GCAAGCTGTCCCTCAGCTTCC------CCAAGCGCCTGCAGGACAACCCCGCGTTTCTCA 1999

Db   2753 GCAAGCTCACGCAGAGCTTCCCGGCCGCCGAGAACCAGCACGCCGTCGCCGGCGACCCGA 2812

Qy   2000 ACTTCCGCACCGAGGCCGGGCGCACGCT---GTACGGCGAGGACGTCTACGTCGGGTACA 2056

Db   2813 ACCGCTACCCGGGCGTCGACAACCAGCAGACGTACAGCGAGGGCATCCACGTCGGGTACC 2872

Qy   2057 GGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACA 2116

Db   2873 GCTGGTTCGACAAGGAGAACGTCAAGCCGCTGTTCCCGTTCGGGCACGGCCTGTCGTACA 2932

Qy   2117 CCACTTTTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGAC---GGCAAGCTGAGCGTGT 2173

Db   2933 CCTCGTTCACGCAGAGCGCCCCGACCGTGGTGCGCACGTCCACGGGCGGCCTGAAGGTCA 2992

Qy   2174 CCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCA 2233

Db   2993 CGGTCACGGTGCGCAACAGCGGGCAGCGCGCGGGCCAGGAGGTCGTCCAGGCGTATCTCG 3052

Qy   2234 AGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGG 2293

Db   3053 GCGCGAGCCCGAAGGTGACGGCTCCGCAGGCGGAGAAGAAGCTCGTGGGCTACACGAAGG 3112

Qy   2294 TCGAACTGCAGCCCGGCGAGACGAAGGCGGTGAC 2327

Db   3113 TCGCGCTCGCGGCGGGCGAGTCGAAGACGGTGAC 3146


RESULT 3
AAA75634
ID   AAA75634 standard; DNA; 2401 BP.
XX
AC   AAA75634;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Nucleotide sequence of ORF11 which encodes 1-beta glucosidase.
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis; ss.
XX
OS   Streptomyces venezuelae.
XX
PN   US6117659-A.
XX
PD   12-SEP-2000.
XX
PF   27-MAY-1999;   99US-0320878.
XX
PR   28-MAY-1998;   98US-0087080.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
PR   20-MAY-1999;   99US-0134990.
PR   30-APR-1997;   97US-0846247.
PR   06-MAY-1998;   98US-0073538.
PR   28-AUG-1998;   98US-0141908.
```

XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX
PT New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT -
XX
PS Disclosure; Columns 39-40; 117pp; English.
XX
CC The present sequence is used to produce the recombinant DNA compounds
CC of the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 2401 BP; 405 A; 859 C; 828 G; 308 T; 1 other;

Query Match 7.2%; Score 179; DB 21; Length 2401;
Best Local Similarity 47.7%; Pred. No. 5.6e-24;
Matches 1018; Conservative 0; Mismatches 966; Indels 150; Gaps 11;



```
Qy   1886 GCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGG 1945
Db   1797 CGGGCCAGGCGGGCGCCGAGGCCACCGCCGCGCTGCTCTACGGTGACGTCAACCCGAGCG 1856

Qy   1946 GCAAGCTGTCCCTCAGCTTCCCC---------AAGCGCCTGCAGGACAACCCCGCGTTTC 1996
Db   1857 GCAAGCTCACGCAGAGCTTCCCGGCCGCCGAGAACCAGCACGCGGTCGCCGGCGACCCGA 1916

Qy   1997 TCAACTTCCGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTCTACGTCGGGTACA 2056
Db   1917 CCAGCTACCCGGGCGTCGACAACCAGCAGACGTACCGCGAGGGCATCCACGTCGGGTACC 1976

Qy   2057 GGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACA 2116
Db   1977 GCTGGTTCGACAAGGAGAACGTCAAGCCGCTGTTCCCGTTCGGGCACGGCCTGTCGTACA 2036

Qy   2117 CCACTTTTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGAC---GGCAAGCTGAGCGTGT 2173
Db   2037 CCTCGTTCACGCAGAGCGCCCCGACCGTCGTGCGTACGTCCACGGGTGGTCTGAAGGTCA 2096

Qy   2174 CCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCA 2233
Db   2097 CGGTCACGGTCCGCAACAGCGGGAAGCGCGCCGGCCAGGAGGTCGTCCAGGCGTACCTCG 2156

Qy   2234 AGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGG 2293
Db   2157 GTGCCAGCCCGAACGTGACGGCTCCGCAGGCGAAGAAGAAGCTCGTGGGCTACACGAAGG 2216

Qy   2294 TCGAACTGCAGCCCGGCGAGACGAAGGCGGTGAC 2327
Db   2217 TCTCGCTCGCCGCGGGCGAGGCGAAGACGGTGAC 2250
```

RESULT 4
AAZ56002
ID AAZ56002 standard; DNA; 2401 BP.
XX
AC AAZ56002;
XX
DT 23-MAR-2000 (first entry)
XX
DE Contig 001 from cosmid pKOS023-27 from Streptomyces venezuelae.
XX
KW Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; contig 001;
KW ketolide; 1 beta glucosidase; antibiotic production; narbomycin;
KW picromycin; ds.
XX
OS Streptomyces venezuelae.
XX
FH Key Location/Qualifiers
FT CDS 80..2389
FT /*tag= a
FT /product= 1_beta_glucosidase
XX
PN WO9961599-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11814.
XX
PR 28-MAY-1998; 98US-0087080.
PR 28-AUG-1998; 98US-0141908.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX
DR WPI; 2000-072618/06.
DR P-PSDB; AAY67212.
XX
PT New recombinant DNA encoding a domain of narbonolide polyketide
PT synthase, for production of ketolide antibiotics -
XX
PS Disclosure; Page 34-35; 98pp; English.
XX
CC This is contig 001 from the recombinant cosmid pKOS023-27 DNA sequence
CC (see AAZ56001) which contains a Streptomyces venezuelae DNA insert. The
CC cosmid contains open reading frames which encode the various modules of
CC the narbonolide polyketide synthase (PKS). The invention relates to
CC recombinant DNA containing a coding sequence for a narbonolide PKS.
CC Polyketides are compounds synthesised from 2-carbon units through a
CC series of condensations and subsequent modifications. Modular PKSs are
CC responsible for the production of many antibiotics including picromycin.
CC The narbonolide PKS consists of a loading module, six extender modules,
CC and two thioester domains. Four proteins make up the narbonolide PKS
CC (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC PICAIII includes extender module 5 and PICAIV includes extender module 6
CC and a type II thioesterase domain. The second type II thioesterase
CC domain is found on the PICB protein. The nucleotide sequences encoding
CC all of these proteins can be isolated in recombinant form from the
CC recombinant cosmid pKOS023-27. Narbonolide is desosaminylated in S.
CC venezuelae to yield narbomycin, the desosaminyl transferase enzyme is
CC required for this conversion, and the desosamine biosynthetic genes are
CC also found in cosmid pKOS023-27. The recombinant DNA of the invention is
CC used to express, in transformed cells, narbonolide (or its derivatives)
CC or other ketolides (particularly hybrids), which may then be converted
CC (e.g. by other enzymes recombinantly expressed in the same hosts) to
CC polyketide antibiotics or their intermediates. The antibiotics are useful
CC in human or veterinary medicine.
XX
SQ Sequence 2401 BP; 405 A; 859 C; 828 G; 308 T; 1 other;

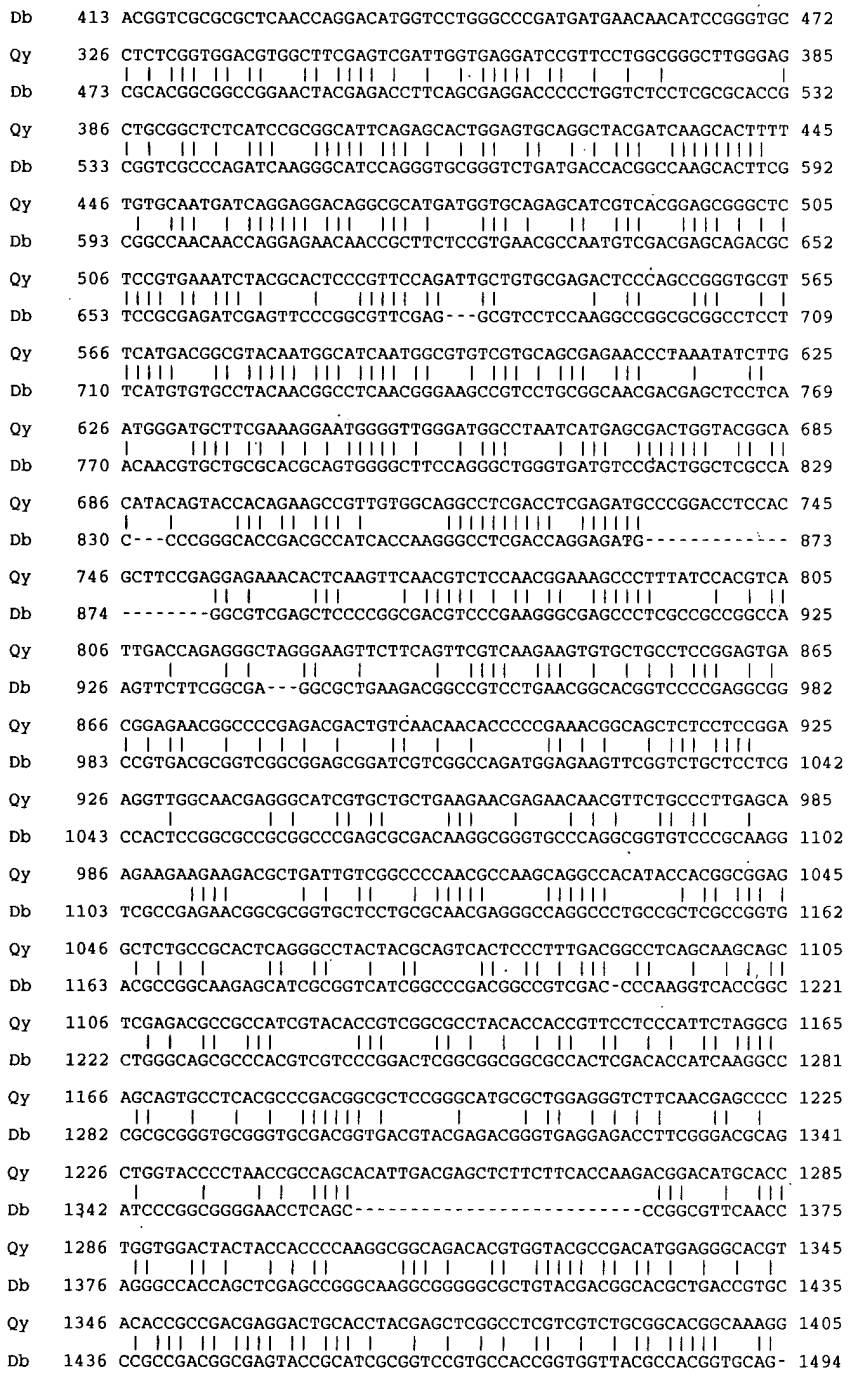
Query Match 7.2%; Score 179; DB 21; Length 2401;
Best Local Similarity 47.7%; Pred. No. 5.6e-24;
Matches 1018; Conservative 0; Mismatches 966; Indels 150; Gaps 11;

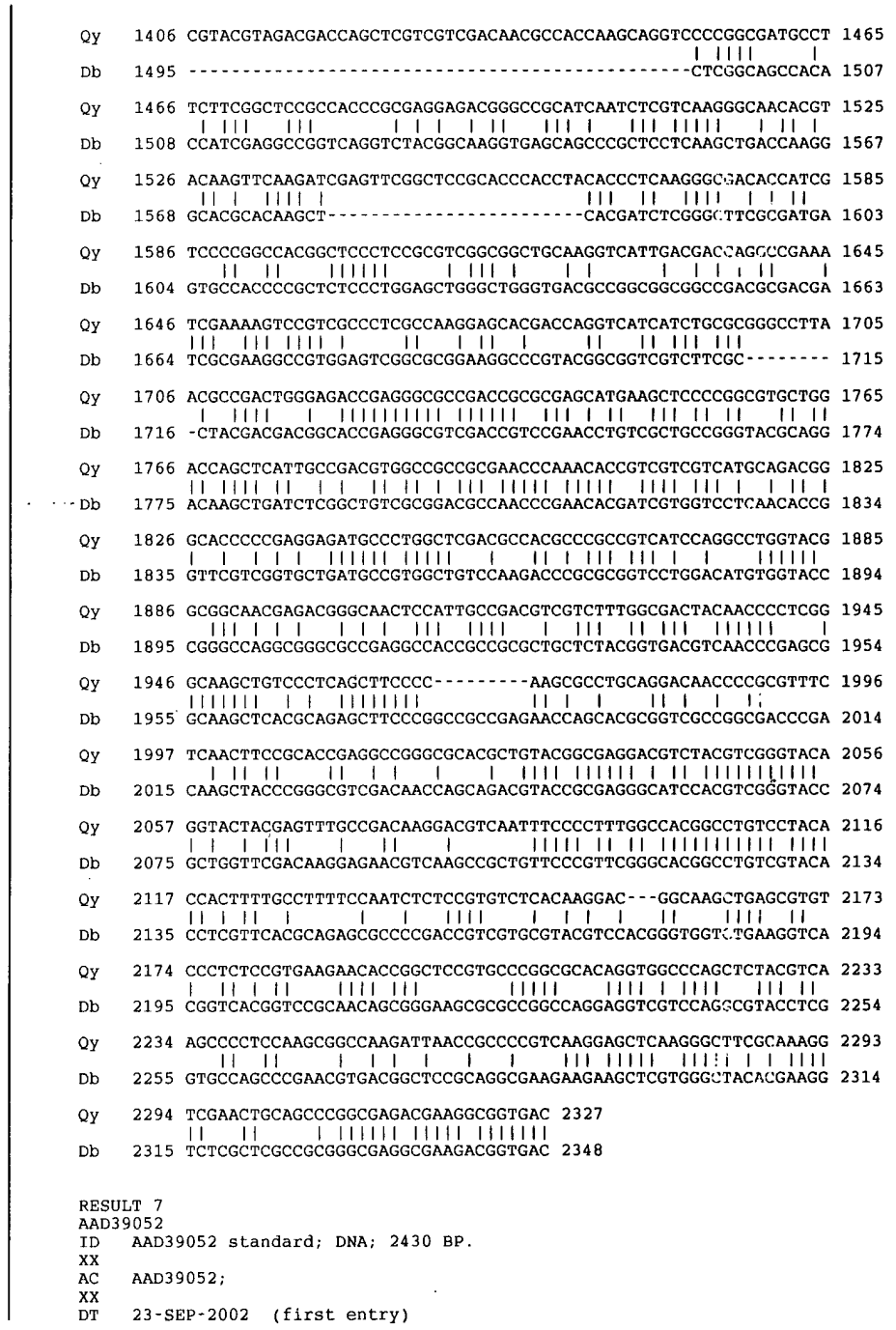
```
Qy    206 CGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAG 265
Db    255 CCCTGGCCAGCACCTTCGACGACACCATGGCCGACAGCTACGGCAAGGTCATGGGCCGCG 314

Qy    266 AGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCC 325
Db    315 ACGGTCGCGCGCTCAACCAGGACATGGTCCTGGGCCCGATGATGAACAACATCCGGGTGC 374

Qy    326 CTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAG 385
Db    375 CGCACGGCGGCCGGAACTACGAGACCTTCAGCGAGGACCCCCTGGTCTCCTCGCGCACCG 434

Qy    386 CTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTT 445
Db    435 CGGTCGCCCAGATCAAGGGCATCCAGGGTGCGGGTCTGATGACCACGGCCAAGCACTTCG 494

Qy    446 TGTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGCATCGTCACGGAGCGGGCTC 505
Db    495 CGGCCAACAACCAGGAGAACAACCGCTTCTCCGTGAACGCCAATGTCGACGAGCAGACGC 554

Qy    506 TCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCCAGCCGGGTGCGT 565
Db    555 TCCGCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCCAAGGCCGGCGCGGGCTCCT 611

Qy    566 TCATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGCGAGAACCCTAAATATCTTG 625
Db    612 TCATGTGTGCCTACAACGGCCTCAACGGGAAGCCGTCCTGCGGCAACGACGAGCTCCTCA 671

Qy    626 ATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCGACTGGTACGGCA 685
Db    672 ACAACGTGCTGCGCACGCAGTGGGGCTTCCAGGGCTGGGTGATGTCCGACTGGCTCGCCA 731

Qy    686 CATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGCCCGGACCTCCAC 745
Db    732 C---CCCGGGCACCGACGCCATCACCAAGGGCCTCGACCAGGAGATG---------- 775
```

```
Qy    746 GCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCTTTATCCACGTCA 805
Db    776 --------GGCGTCGAGCTCCCCGGCGACGTCCCGAAGGGCGAGCCCTCGCCGCCGGCCA 827

Qy    806 TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTGCTGCCTCCGGAGTGA 865
Db    828 AGTTCTTCGGCGA---GGCGCTGAAGACGGCCGTCCTGAACGGCACGGTCCCCGAGGCGG 884

Qy    866 CGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCCGAAACGGCAGCTCTCCTCCGGA 925
Db    885 CCGTGACGCGGTCGGCGGAGCGGATCGTCGGCCAGATGGAGAAGTTCGGTCTGCTCCTCG 944

Qy    926 AGGTTGGCAACGAGGGCATCGTGCTGCTGAAGAACGAGAACAACGTTCTGCCCTTGAGCA 985
Db    945 CCACTCCGGCGCCGCGGCCCGAGCGCGACAAGGCGGGTGCCCAGGCGGTGTCCCGCAAGG 1004

Qy    986 AGAAGAAGAAGACGCTGATTGTCGGCCCCAACGCCAAGCAGGCCACATACCACGGCGGAG 1045
Db   1005 TCGCCGAGAACGGCGCGGTGCTCCTGCGCAACGAGGGCCAGGCCCTGCCGCTCGCCGGTG 1064

Qy   1046 GCTCTGCCGCACTCAGGGCCTACTACGCAGTCACTCCCTTTGACGGCCTCAGCAAGCAGC 1105
Db   1065 ACGCCGGCAAGAGCATCGCGGTCATCGGCCCGACGGCCGTCGAC-CCCAAGGTCACCGGC 1123

Qy   1106 TCGAGACGCCGCCATCGTACACCGTCGGCGCCTACACCACCGTTCCTCCCATTCTAGGCG 1165
Db   1124 CTGGGCAGCGCCCACGTCGTCCCGGACTCGGCGGCGGCGCCACTCGACACCATCAAGGCC 1183

Qy   1166 AGCAGTGCCTCACGCCCGACGGCGCTCCGGGCATGCGCTGGAGGGTCTTCAACGAGCCCC 1225
Db   1184 CGCGCGGGTGCGGGTGCGACGGTGACGTACGAGACGGGTGAGGAGACCTTCGGGACGCAG 1243

Qy   1226 CTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCTTCACCAAGACGGACATGCACC 1285
Db   1244 ATCCCGGCGGGGAACCTCAGC------------------------CCGGCGTTCAACC 1277

Qy   1286 TGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGGTACGCCGACATGGAGGGCACGT 1345
Db   1278 AGGGCCACCAGCTCGAGCCGGGCAAGGCGGGGGCGCTGTACGACGGCACGCTGACCGTGC 1337

Qy   1346 ACACCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGTCGTCTGCGGCACGGCAAAGG 1405
Db   1338 CCGCCGACGGCGAGTACCGCATCGCGGTCCGTGCCACCGGTGGTTACGCCACGGTGCAG- 1396

Qy   1406 CGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCCCCGGCGATGCCT 1465
Db   1397 ----------------------------------------------CTCGGCAGCCACA 1409

Qy   1466 TCTTCGGCTCCGCCACCCGCGAGGAGACGGGCCGCATCAATCTCGTCAAGGGCAACACGT 1525
Db   1410 CCATCGAGGCCGGTCAGGTCTACGGCAAGGTGAGCAGCCCGCTCCTCAAGCTGACCAAGG 1469

Qy   1526 ACAAGTTCAAGATCGAGTTCGGCTCCGCACCCACCTACACCCTCAAGGGCGACACCATCG 1585
Db   1470 GCACGCACAAGCT-----------------------CACGATCTCGGGCTTCGCGATGA 1505

Qy   1586 TCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAAA 1645
Db   1506 GTGCCACCCCGCTCTCCCTGGAGCTGGGCTGGGTNACGCCGGCGGCGGCCGACGCGACGA 1565

Qy   1646 TCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTTA 1705
Db   1566 TCGCGAAGGCCGTGGAGTCGGCGCGGAAGGCCCGTACGGCGGTCGTCTTCGC-------- 1617

Qy   1706 ACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGG 1765
Db   1618 -CTACGACGACGGCACCGAGGGCGTCGACCGTCCGAACCTGTCGCTGCCGGGTACGCAGG 1676

Qy   1766 ACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGG 1825
Db   1677 ACAAGCTGATCTCGGCTGTCGCGGACGCCAACCCGAACACGATCGTGGTCCTCAACACCG 1736

Qy   1826 GCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACG 1885
Db   1737 GTTCGTCGGTGCTGATGCCGTGGCTGTCCAAGACCCGCGCGGTCCTGGACATGTGGTACC 1796

Qy   1886 GCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGG 1945
Db   1797 CGGGCCAGGCGGGCGCCGAGGCCACCGCCGCGCTGCTCTACGGTGACGTCAACCCGAGCG 1856

Qy   1946 GCAAGCTGTCCCTCAGCTTCCCC---------AAGCGCCTGCAGGACAACCCCGCGTTTC 1996
Db   1857 GCAAGCTCACGCAGAGCTTCCCGGCCGCCGAGAACCAGCACGCGGTCGCCGGCGACCCGA 1916

Qy   1997 TCAACTTCCGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTCTACGTCGGGTACA 2056
Db   1917 CCAGCTACCCGGGCGTCGACAACCAGCAGACGTACCGCGAGGGCATCCACGTCGGGTACC 1976

Qy   2057 GGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACA 2116
Db   1977 GCTGGTTCGACAAGGAGAACGTCAAGCCGCTGTTCCCGTTCGGGCACGGCCTGTCGTACA 2036

Qy   2117 CCACTTTTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGAC---GGCAAGCTGACGTGT 2173
Db   2037 CCTCGTTCACGCAGAGCGCCCCGACCGTCGTGCGTACGTCCACGGGTGGTCTGAAGGTCA 2096

Qy   2174 CCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCA 2233
Db   2097 CGGTCACGGTCCGCAACAGCGGGAAGCGCGCCGGCCAGGAGGTCGTCCAGGCGTACCTCG 2156

Qy   2234 AGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGG 2293
Db   2157 GTGCCAGCCCGAACGTGACGGCTCCGCAGGCGAAGAAGAAGCTCGTGGGCTACACGAAGG 2216

Qy   2294 TCGAACTGCAGCCCGGCGAGACGAAGGCGGTGAC 2327
Db   2217 TCTCGCTCGCCGCGGGCGAGGCGAAGACGGTGAC 2250
```

RESULT 5
AAT04785
ID AAT04785 standard; DNA; 1145 BP.
XX
AC AAT04785;
XX
DT 17-JAN-1996 (first entry)
XX
DE Cellobiase gene fragment O1-O5.
XX
KW Cellobiase; beta-glucosidase; cellulose; waste-disposal; ethanol;
KW Escherichia coli; Saccharomyces cerevisiae; ds.
XX
OS Cellulomonas biazotea.
XX
FH Key Location/Qualifiers
FT CDS 331..1145
FT /*tag= a
XX
PN GB2289050-A.
XX
PD 08-NOV-1995.
XX
PF 05-MAY-1995; 95GB-0009237.
XX
PR 06-MAY-1994; 94GB-0009030.
XX
PA (UYHK-) UNIV HONG KONG.
XX
PI Chan WK, Wong WK;
XX
DR WPI; 1995-368626/48.
XX
PT New cellobiase from Cellulomonas biazotea and related nucleic acid
PT - used to degrade cellulosic waste, esp. to ethanol in conjunction
PT with yeast glucanase(s)
XX

```
PS   Disclosure; Fig.14; 41pp; English.
XX
CC   The 0.75-kb PstI and the 3.05-kb Pst-NdeI fragments of C. biazotea
CC   ATCC 486 chromosomal DNA were sequenced using the dideoxy method.  8
CC   Sections of the coding strand were identified, the first 5 of which
CC   were contiguous (given in AAT04785) and included a putative start
CC   codon.  The sequences of the other 3, non-contiguous, downstream
CC   fragments are given in AAQ73991-93.  Cloning of the gene region allowed
CC   extracellular prodn. of recombinant cellobiase in E. coli and S.
CC   cerevisiae.
XX
SQ   Sequence 1145 BP; 152 A; 444 C; 391 G; 158 T; 0 other;

  Query Match             7.1%;  Score 177.4;  DB 16;  Length 1145;
  Best Local Similarity   57.2%;  Pred. No. 1e-23;
  Matches  362;  Conservative    0;  Mismatches  266;  Indels    5;  Gaps    2;

Qy     109 GGAGTCCCCTCTCTCCGCTTTACAGATGGCCCCAACGGCGTAAGAGGGACCAAGTTCTTC 168
           ||  ||  |    ||||| |  | || ||||| | ||||||  | ||   |||||||| |
Db     448 GGCCTCGGCGAGGTCCGCCTGTCCGACGGCCCGACCGGCGTGCGCGGCCTCAAGTTCTCC 507

Qy     169 AATGGCGTCCCTGCGGCCTGCTTCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAA 228
             ||   | | |  ||    |||||   ||  ||   |||||  ||| | |  | | |
Db     508 GGCGGGCGCACCGTCGCGCTGTTCCCCAACGCGACCCTGCTCGCGTCCGCGTGGAGCGAG 567

Qy     229 ACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCAT 288
               |      ||||  ||    | || | | | | |||||| ||||  ||       ||
Db     568 GAGTCGACGACCGAGGTCGGCCGGCTGCTCGCCGAGGAGGCCCTCGCGCAGCAGATCCAC 627

Qy     289 GTGATCCTCGGCCCGACTATCAACATGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAG 348
           ||  | |||||||| || |||||| |||| ||||||   ||||| |||||    ||||||
Db     628 GTCGTGCTCGGCCCCACGATCAACCTGCACCGCTCCGTGCTCGGCGGACGCCTGTTCGAG 687

Qy     349 TCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATT 408
             |      ||||| ||| | ||  | |||  |   || || ||   | | ||||||  |
Db     688 GCCTACTCCGAGGACCCGCTGCTCACCGGCCGGCTCGCCGCCGCCTACGTGCGCGGCCTG 747

Qy     409 CAGAGCACTGGAGT-GCAGGCTACGATCAAGCACTTTTTGTGCAATGATCAGGAGGACAG 467
           |||  |   || || |  | || |    | |||| |  |   ||| ||   ||||  |
Db     748 CAGGACCTCGGCGTCGGCGCCTGCCTCAAGGCACCTCGTCGCCAACGAGTCGGAGACCGA 807

Qy     468 GCGCATGATGGTGCAGAGCATCGTCACGGAGCGGGCTCTCCGTGAAATCTACGCACTCCC 527
           |||||  |   || |   | |||||       | | ||||| ||  |||||   || ||
Db     808 GCGCAACACCATGAACTCCGTCGTCGACCCCGCGACGCTCCGCGAGCTCTACCTGCTGCC 867

Qy     528 GTTCCAGATTGCTGTGCGAGACTCCCAGCCGGGTGCGTTCATGACGGCGTACAATGGCAT 587
           |||| |||| || ||    || ||| | ||| |  |  ||||| | |||||||| | | |
Db     868 GTTCGAGATCGCCGTCGACGAGTCCGACCCGTGGTCCGTCATGGCCGCGTACAACGACGT 927

Qy     588 CAATGGCGTGTCGTGCAGCGAGAACCCTAAATATCTTGATGGGATGCTTCGAAAGGAATG 647
           ||| |||||  |    | |||| |||   |     |  | | | | ||       || ||
Db     928 CAACGGCGTCCCCGCGACCGAGCACCACCACGTCGTCAACGAGGTCCTCAAGGGCGAGTG 987

Qy     648 GGGTTGGGATGGCCTAATCATGAGCGACTGGTACGGCACATACAGTACCACAGAAGCCGT 707
           ||| |     |||||  |||||  |||||||| ||  ||    |   || ||   ||||
Db     988 GGGCTACACCGGCCTCGTCATGTCCGACTGGTTCGCGACCCGGACCGCCGCACCCGCCGC 1047

Qy     708 TGTGGCAGGCCTCGACCTCGAGATGCCCGGACC 740
           |      ||||||||||||  |||||||| ||
Db    1048 CG----CCGCCTCGACCTCGTCATGCCCGGGCC 1076


RESULT 6
AAZ87294
ID   AAZ87294 standard; DNA; 2430 BP.
XX
AC   AAZ87294;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide beta-glycosidase gene desR, SEQ ID NO:23.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2340
FT                   /*tag=  a
FT                   /product= "DesR"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77189.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 3; Page 369-370; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAZ87286-Z87294 represent desosamine biosynthetic genes from
CC   Streptomyces venezuelae ATCC 15439, which encode proteins
CC   AAY77181-Y77189.
XX
SQ   Sequence 2430 BP; 407 A; 878 C; 834 G; 311 T; 0 other;

  Query Match             7.0%;  Score 175.2;  DB 21;  Length 2430;
  Best Local Similarity   47.6%;  Pred. No. 2.8e-23;
  Matches 1016;  Conservative    0;  Mismatches  968;  Indels  150;  Gaps   11;

Qy     206 CGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAG 265
           | || |    ||| ||| || | ||  ||  |||       || ||| | ||||||   |
Db     353 CCCTGGCCAGCACCTTCGACGACACCATGGCCGACAGCTACGGCAAGGTCATGGGCCGCG 412

Qy     266 AGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCC 325
           | |    |||    |    | |  || |||| |||||||  || ||||     ||    |
```

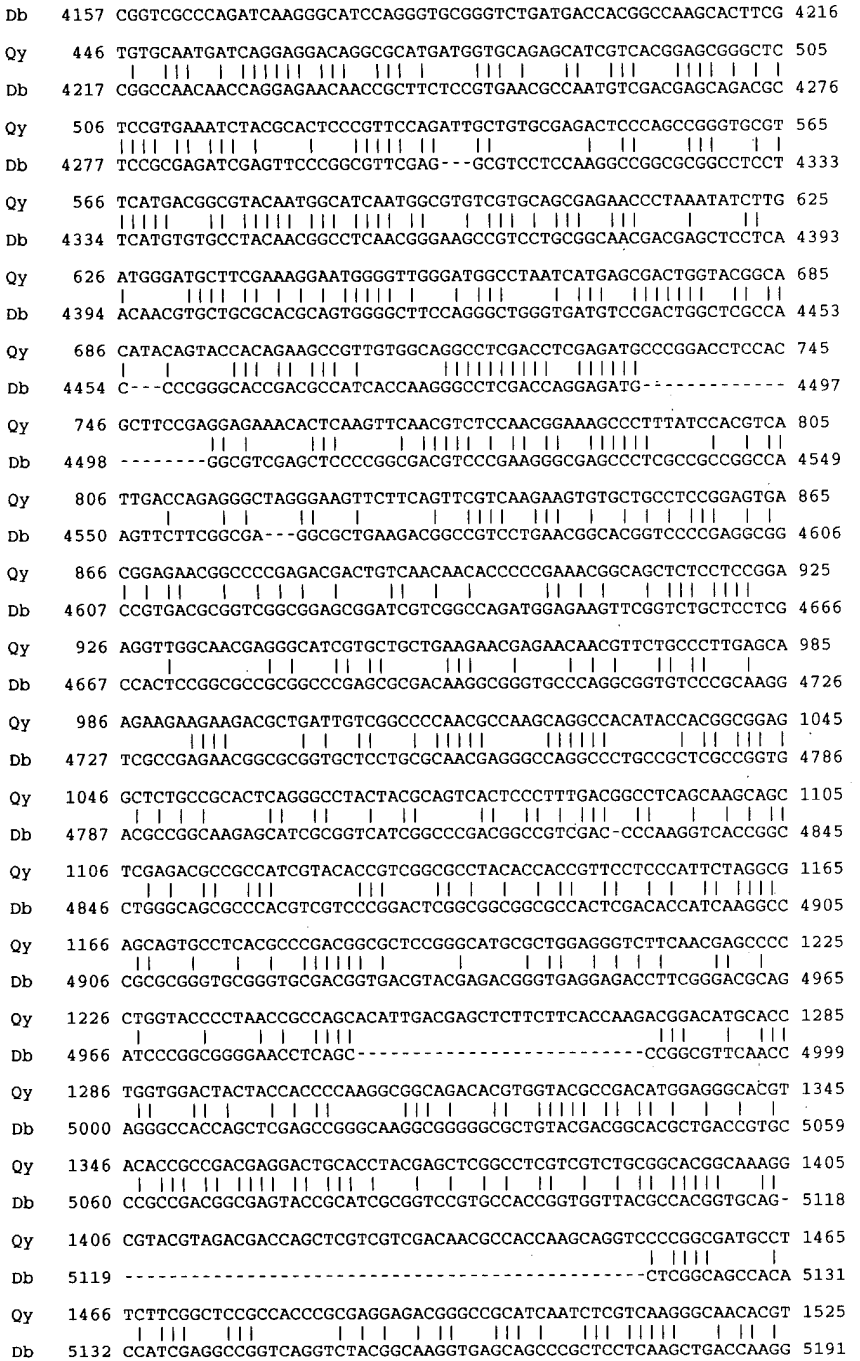


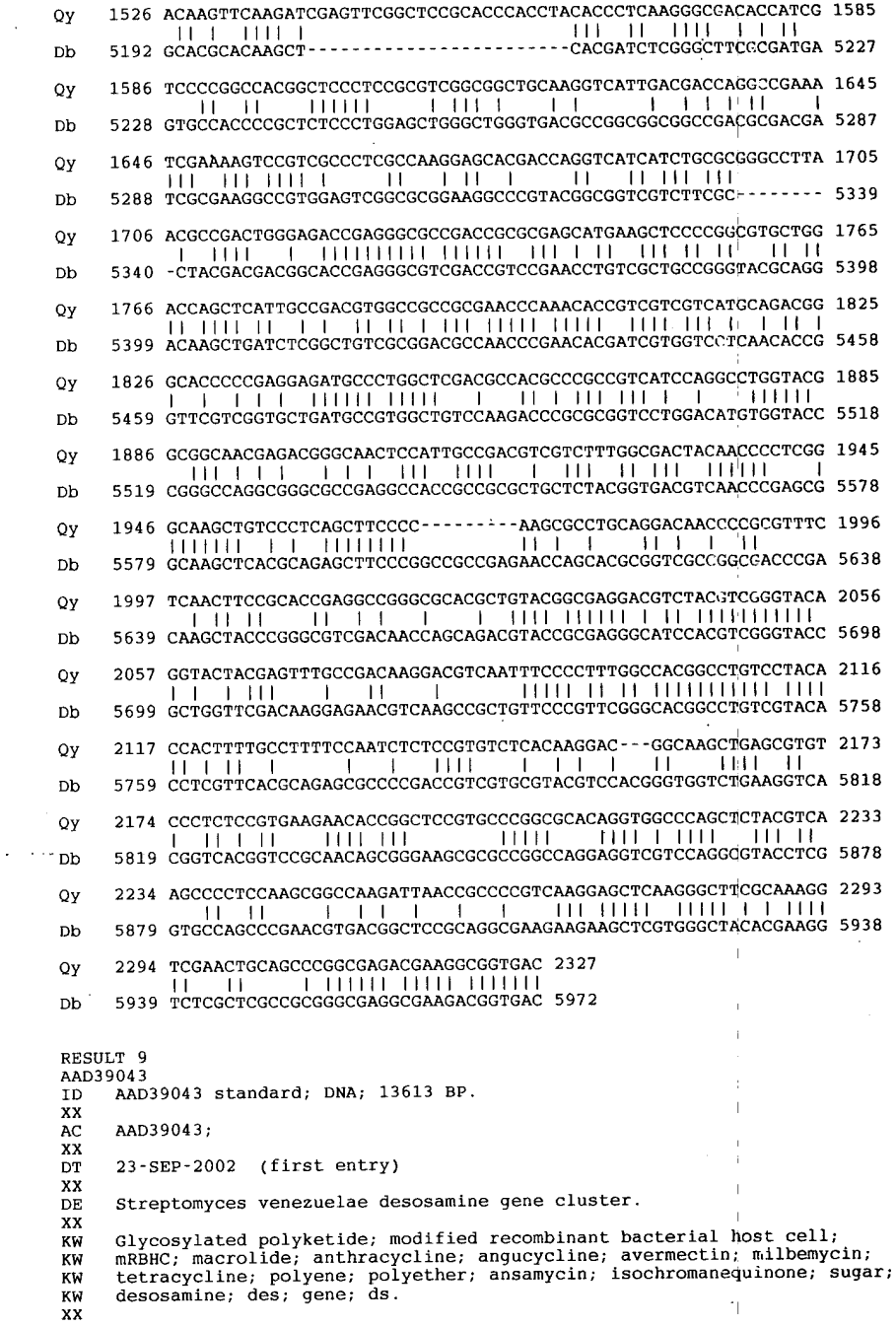


RESULT 7
AAD39052
ID AAD39052 standard; DNA; 2430 BP.
XX
AC AAD39052;
XX
DT 23-SEP-2002 (first entry)

```
XX
DE   Streptomyces venezuelae DesR gene.
XX
KW   Glycosylated polyketide; modified recombinant bacterial host cell;
KW   mRBHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;
KW   tetracycline; polyene; polyether; ansamycin; isochromanequinone; sugar;
KW   desosamine; DesR; glucosidase; enzyme; gene; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2430
FT                   /*tag=  a
FT                   /product= "Streptomyces venezuelae DesR protein"
FT                   /transl_except= (pos:1..3, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200229035-A2.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31255.
XX
PR   05-OCT-2000; 2000US-238185P.
XX
PA   (MINU ) UNIV MINNESOTA.
PA   (LIUH/) LIU H.
PA   (SHER/) SHERMAN D H.
PA   (ZHAO/) ZHAO L.
XX
PI   Liu H,  Sherman DH,  Zhao L;
XX
DR   WPI; 2002-405171/43.
DR   P-PSDB; AAE24237.
XX
PT   Modified recombinant bacterial host cells in which the expression and
PT   activity of nucleic acids encoding sugar biosynthetic enzymes has been
PT   altered, useful for producing metabolites with altered sugar structures
PT   -
XX
PS   Disclosure; Page 170-171; 174pp; English.
XX
CC   The invention provides a method to alter the sugar structure diversity
CC   for a particular metabolite via the recruitment and collaborative action
CC   of sugar genes from a variety of sugar biosynthetic pathways to yield a
CC   metabolite comprising a non-natural sugar, e.g., a novel glycosylated
CC   polyketide. The invention also relates to a modified recombinant
CC   bacterial host cell (mRBHC) in which the expression and activity of
CC   nucleic acids encoding sugar biosynthetic enzymes has been altered.
CC   The mRBHCs may be cultured to produce the modified sugar products,
CC   e.g. a macrolide, anthracycline, angucycline, avermectin, milbemycin,
CC   tetracycline, polyene, polyether, ansamycin or isochromanequinone.
CC   The present sequence is Streptomyces venezuelae sugar (desosamine)
CC   biosynthetic gene cluster DesR (glucosidase) gene.
XX
SQ   Sequence 2430 BP; 407 A; 878 C; 834 G; 311 T; 0 other;

  Query Match             7.0%;  Score 175.2;  DB 24;  Length 2430;
  Best Local Similarity   47.6%;  Pred. No. 2.8e-23;
  Matches 1016;  Conservative    0;  Mismatches  968;  Indels  150;  Gaps   11;

Qy    206 CGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAG 265
Db    353 CCCTGGCCAGCACCTTCGACGACACCATGGCCGACAGCTACGGCAAGGTCATGGGCCGCG 412

Qy    266 AGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCC 325
Db    413 ACGGTCGCGCGCTCAACCAGGACATGGTCCTGGGCCCGATGATGAACAACATCCGGGTGC 472

Qy    326 CTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAG 385
Db    473 CGCACGGCGGCCGGAACTACGAGACCTTCAGCGAGGACCCCCTGGTCTCCTCGCGCACCG 532

Qy    386 CTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTT 445
Db    533 CGGTCGCCCAGATCAAGGGCATCCAGGGTGCGGGTCTGATGACCACGGCCAAGCACTTCG 592

Qy    446 TGTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGCATCGTCACGGAGCGGGCTC 505
Db    593 CGGCCAACAACCAGGAGAACAACCGCTTCTCCGTGAACGCCAATGTCGACGAGCAGACGC 652

Qy    506 TCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCCAGCCGGGTGCGT 565
Db    653 TCCGCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCCAAGGCCGGCGCGGCCTCCT 709

Qy    566 TCATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGCGAGAACCCTAAATATCTTG 625
Db    710 TCATGTGTGCCTACAACGGCCTCAACGGGAAGCCGTCCTGCGGCAACGACGAGCTCCTCA 769

Qy    626 ATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCGACTGGTACGGCA 685
Db    770 ACAACGTGCTGCGCACGCAGTGGGGCTTCCAGGGCTGGGTGATGTCCGACTGGCTCGCCA 829

Qy    686 CATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGCCCGGACCTCCAC 745
Db    830 C---CCCGGGCACCGACGCCATCACCAAGGGCCTCGACCAGGAGATG------------- 873

Qy    746 GCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCTTTATCCACGTCA 805
Db    874 --------GGCGTCGAGCTCCCCGGCGACGTCCCGAAGGGCGAGCCCTCGCCGCCGGCCA 925

Qy    806 TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTGCTGCCTCCGGAGTGA 865
Db    926 AGTTCTTCGGCGA---GGCGCTGAAGACGGCCGTCCTGAACGGCACGGTCCCCGAGGCGG 982

Qy    866 CGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCCGAAACGGCAGCTCTCCTCCGGA 925
Db    983 CCGTGACGCGGTCGGCGGAGCGGATCGTCGGCCAGATGGAGAAGTTCGGTCTGCTCCTCG 1042

Qy    926 AGGTTGGCAACGAGGGCATCGTGCTGCTGAAGAACGAGAACAACGTTCTGCCCTTGAGCA 985
Db   1043 CCACTCCGGCGCCGCGGCCCGAGCGCGACAAGGCGGGTGCCCAGGCGGTGTCCCGCAAGG 1102

Qy    986 AGAAGAAGAAGACGCTGATTGTCGGCCCCAACGCCAAGCAGGCCACATACCACGGCGGAG 1045
Db   1103 TCGCCGAGAACGGCGCGGTGCTCCTGCGCAACGAGGGCCAGGCCCTGCCGCTCGCCGGTG 1162

Qy   1046 GCTCTGCCGCACTCAGGGCCTACTACGCAGTCACTCCCTTTGACGGCCTCAGCAAGCAGC 1105
Db   1163 ACGCCGGCAAGAGCATCGCGGTCATCGGCCCGACGGCCGTCGAC-CCCAAGGTCACCGGC 1221

Qy   1106 TCGAGACGCCGCCATCGTACACCGTCGGCGCCTACACCACCGTTCCTCCCATTCTAGGCG 1165
Db   1222 CTGGGCAGCGCCCACGTCGTCCCGGACTCGGCGGCGGCGCCACTCGACACCATCAAGGCC 1281

Qy   1166 AGCAGTGCCTCACGCCCGACGGCGCTCCGGGCATGCGCTGGAGGGTCTTCAACGAGCCCC 1225
Db   1282 CGCGCGGGTGCGGGTGCGACGGTGACGTACGAGACGGGTGAGGAGACCTTCGGGACGCAG 1341

Qy   1226 CTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCTTCACCAAGACGGACATGCACC 1285
Db   1342 ATCCCGGCGGGGAACCTCAGC------------------------CCGGCGTTCAACC 1375

Qy   1286 TGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGGTACGCCGACATGGAGGGCACGT 1345
Db   1376 AGGGCCACCAGCTCGAGCCGGGCAAGGCGGGGGCGCTGTACGACGGCACGCTGACCGTGC 1435

Qy   1346 ACACCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGTCGTCTGCGGCACGGCAAAGG 1405
Db   1436 CCGCCGACGGCGAGTACCGCATCGCGGTCCGTGCCACCGGTGGTTACGCCACGGTGCAG- 1494

Qy   1406 CGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCCCCGGCGATGCCT 1465
Db   1495 ----------------------------------------------CTCGGCAGCCACA 1507
```

```
Qy   1466 TCTTCGGCTCCGCCACCCGCGAGGAGACGGGCCGCATCAATCTCGTCAAGGGCAACACGT 1525
          | |||   |||       | | |  | ||   ||| |   ||| |||||   | || |
Db   1508 CCATCGAGGCCGGTCAGGTCTACGGCAAGGTGAGCAGCCCGCTCCTCAAGCTGACCAAGG 1567

Qy   1526 ACAAGTTCAAGATCGAGTTCGGCTCCGCACCCACCTACACCCTCAAGGGCGACACCATCG 1585
          || |  |||| |                          |||  ||  ||||  | | ||
Db   1568 GCACGCACAAGCT----------------------CACGATCTCGGGCTTCGCGATGA 1603

Qy   1586 TCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAAA 1645
            ||  ||    ||||||     | ||| |    | |       |  | | | ||    |
Db   1604 GTGCCACCCCGCTCTCCCTGGAGCTGGGCTGGGTGACGCCGGCGGCGGCCGACGCGACGA 1663

Qy   1646 TCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTTA 1705
          |||  ||| |||| |     ||   | ||  |     ||   || |||  |||
Db   1664 TCGCGAAGGCCGTGGAGTCGGCGCGGAAGGCCCGTACGGCGGTCGTCTTCGC-------- 1715

Qy   1706 ACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGG 1765
          |  ||||   |  |||||||||| ||||||  ||| | ||  ||| || ||   || ||
Db   1716 -CTACGACGACGGCACCGAGGGCGTCGACCGTCCGAACCTGTCGCTGCCGGGTACGCAGG 1774

Qy   1766 ACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGG 1825
          || |||| ||  | |  || || | ||| ||||| |||||  |||| ||| |  | || |
Db   1775 ACAAGCTGATCTCGGCTGTCGCGGACGCCAACCCGAACACGATCGTGGTCCTCAACACCG 1834

Qy   1826 GCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACG 1885
          |  |  | | |  |||||| |||||    |   || | ||| ||| |  |    ||||||
Db   1835 GTTCGTCGGTGCTGATGCCGTGGCTGTCCAAGACCCGCGCGGTCCTGGACATGTGGTACC 1894

Qy   1886 GCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGG 1945
            ||| | | |   | | |  |||  ||||   |  |||  || |||  ||||||  |
Db   1895 CGGGCCAGGCGGGCGCCGAGGCCACCGCCGCGCTGCTCTACGGTGACGTCAACCCGAGCG 1954

Qy   1946 GCAAGCTGTCCCTCAGCTTCCCC---------AAGCGCCTGCAGGACAACCCCGCGTTTC 1996
          |||||||  | |  ||||||||          || |  |    || |  |  ||
Db   1955 GCAAGCTCACGCAGAGCTTCCCGGCCGCCGAGAACCAGCACGCGGTCGCCGGCGACCCGA 2014

Qy   1997 TCAACTTCCGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTCTACGTCGGGTACA 2056
          | || ||   ||  | |   |   |  |||| |||||| | || ||||||||||
Db   2015 CAAGCTACCCGGGCGTCGACAACCAGCAGACGTACCGCGAGGGCATCCACGTCGGGTACC 2074

Qy   2057 GGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACA 2116
          | |  | |||    |   ||    |       ||||| || || ||||||||||| ||||
Db   2075 GCTGGTTCGACAAGGAGAACGTCAAGCCGCTGTTCCCGTTCGGGCACGGCCTGTCGTACA 2134

Qy   2117 CCACTTTTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGAC---GGCAAGCTGAGCGTGT 2173
          || | ||  |      |   |  ||||    | | |  |    ||    ||||  ||
Db   2135 CCTCGTTCACGCAGAGCGCCCCGACCGTCGTGCGTACGTCCACGGGTGGTCTGAAGGTCA 2194

Qy   2174 CCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCA 2233
          |  || | ||    |||| |||        |||||    |||| | ||||   ||| ||
Db   2195 CGGTCACGGTCCGCAACAGCGGGAAGCGCGCCGGCCAGGAGGTCGTCCAGGCGTACCTCG 2254

Qy   2234 AGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGG 2293
            ||  ||     |  | |  |   |   |    ||| |||||  ||||| | | ||||
Db   2255 GTGCCAGCCCGAACGTGACGGCTCCGCAGGCGAAGAAGAAGCTCGTGGGCTACACGAAGG 2314

Qy   2294 TCGAACTGCAGCCCGGCGAGACGAAGGCGGTGAC 2327
          ||   ||     | |||||| ||||| |||||||
Db   2315 TCTCGCTCGCCGCGGGCGAGGCGAAGACGGTGAC 2348


RESULT 8
AAZ87284
ID   AAZ87284 standard; DNA; 12441 BP.
XX
AC   AAZ87284;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae desosamine biosynthetic gene cluster pikB, SEQ ID NO:3.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77179.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 2; Page 281-287; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the desosamine
CC   biosynthetic gene cluster from Streptomyces venezuelae ATCC 15439.
XX
SQ   Sequence 12441 BP; 1704 A; 4294 C; 4686 G; 1757 T; 0 other;

  Query Match             7.0%;  Score 175.2;  DB 21;  Length 12441;
  Best Local Similarity   47.6%;  Pred. No. 3.1e-23;
  Matches 1016;  Conservative    0;  Mismatches  968;  Indels  150;  Gaps   11;

Qy    206 CGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAG 265
          | || |    ||| ||| || | ||  ||  |||       || ||| | ||||||   |
Db   3977 CCCTGGCCAGCACCTTCGACGACACCATGGCCGACAGCTACGGCAAGGTCATGGGCCGCG 4036

Qy    266 AGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCC 325
          | |    |||    |    | |  || |||| |||||||  || ||||     ||    |
Db   4037 ACGGTCGCGCGCTCAACCAGGACATGGTCCTGGGCCCGATGATGAACAACATCCGGGTGC 4096

Qy    326 CTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAG 385
          | | ||| || ||   || |||| |  |  | ||||| ||  |  | |  |          |
Db   4097 CGCACGGCGGCCGGAACTACGAGACCTTCAGCGAGGACCCCCTGGTCTCCTCGCGCACCG 4156

Qy    386 CTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTT 445
          | |  || |  |||   ||||| ||| |  | ||  ||  | | |||  |||||||||
```





RESULT 9
AAD39043
ID AAD39043 standard; DNA; 13613 BP.
XX
AC AAD39043;
XX
DT 23-SEP-2002 (first entry)
XX
DE Streptomyces venezuelae desosamine gene cluster.
XX
KW Glycosylated polyketide; modified recombinant bacterial host cell;
KW mRBHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;
KW tetracycline; polyene; polyether; ansamycin; isochromanequinone; sugar;
KW desosamine; des; gene; ds.
XX

```
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..809
FT                   /*tag= a
FT                   /product= "Streptomyces venezuelae Des gene cluster
FT                   encoded protein #1"
FT   CDS             806..2014
FT                   /*tag= b
FT                   /product= "Streptomyces venezuelae Des VIII protein"
FT                   /transl_except= (pos:806..808, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             2162..3442
FT                   /*tag= c
FT                   /product= "Streptomyces venezuelae Des VII protein"
FT   CDS             3535..4245
FT                   /*tag= d
FT                   /product= "Streptomyces venezuelae Des VI protein"
FT                   /transl_except= (pos:3532..3534, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             4312..6741
FT                   /*tag= e
FT                   /product= "Streptomyces venezuelae Des R protein"
FT                   /transl_except= (pos:4312..4314, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement (7969..6829)
FT                   /*tag= f
FT                   /product= "Streptomyces venezuelae Des V protein"
FT                   /transl_except= (pos:7969..7967, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement (8979..7966)
FT                   /*tag= g
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT                   /transl_except= (pos:8979..8977, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
FT   CDS             complement (9111..9989)
FT                   /*tag= h
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT   CDS             complement (11639..10182)
FT                   /*tag= i
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT   CDS             complement (12883..11636)
FT                   /*tag= j
FT                   /product= "Streptomyces venezuelae Des IV protein"
FT                   /transl_except= (pos:12881..12883, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200229035-A2.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31255.
XX
PR   05-OCT-2000; 2000US-238185P.
XX
PA   (MINU ) UNIV MINNESOTA.
PA   (LIUH/) LIU H.
PA   (SHER/) SHERMAN D H.
PA   (ZHAO/) ZHAO L.
XX
PI   Liu H,  Sherman DH,  Zhao L;
XX
DR   WPI; 2002-405171/43.
DR   P-PSDB; AAE24228, AAE24229, AAE24230, AAE24231, AAE24232, AAE24233,
DR   AAE24234, AAE24235, AAE24236, AAE24237, AAE24347.
XX
PT   Modified recombinant bacterial host cells in which the expression and
PT   activity of nucleic acids encoding sugar biosynthetic enzymes has been
PT   altered, useful for producing metabolites with altered sugar structures
PT   -
XX
PS   Disclosure; Fig 8; 174pp; English.
XX
CC   The invention provides a method to alter the sugar structure diversity
CC   for a particular metabolite via the recruitment and collaborative action
CC   of sugar genes from a variety of sugar biosynthetic pathways to yield a
CC   metabolite comprising a non-natural sugar, e.g., a novel glycosylated
CC   polyketide. The invention also relates to a modified recombinant
CC   bacterial host cell (mRBHC) in which the expression and activity of
CC   nucleic acids encoding sugar biosynthetic enzymes has been altered.
CC   The mRBHCs may be cultured to produce the modified sugar products,
CC   e.g. a macrolide, anthracycline, angucycline, avermectin, milbemycin,
CC   tetracycline, polyene, polyether, ansamycin or isochromanequinone.
CC   The present sequence is Streptomyces venezuelae sugar (desosamine)
CC   biosynthetic gene cluster.
XX
SQ   Sequence 13613 BP; 1858 A; 4732 C; 5093 G; 1930 T; 0 other;

  Query Match             7.0%;  Score 175.2;  DB 24;  Length 13613;
  Best Local Similarity   47.6%;  Pred. No. 3.1e-23;
  Matches 1016;  Conservative    0;  Mismatches  968;  Indels  150;  Gaps   11;

Qy     206 CGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAG 265
           | || |    ||| ||| || | ||  ||  |||        || ||| | ||||||   |
Db    4664 CCCTGGCCAGCACCTTCGACGACACCATGGCCGACAGCTACGGCAAGGTCATGGGCCGCG 4723

Qy     266 AGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCC 325
           | |    |||     |    | |  || |||| |||||||  || ||||    ||    |
Db    4724 ACGGTCGCGCGCTCAACCAGGACATGGTCCTGGGCCCGATGATGAACAACATCCGGGTGC 4783

Qy     326 CTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAG 385
           | | ||| || ||   || |||| |  |  | |||||  ||  |  |  |          |
Db    4784 CGCACGGCGGCCGGAACTACGAGACCTTCAGCGAGGACCCCCTGGTCTCCTCGCGCACCG 4843

Qy     386 CTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTT 445
           | |  || |  |||   ||||| ||| |  | ||  ||  | | |||  |||||||||
Db    4844 CGGTCGCCCAGATCAAGGGCATCCAGGGTGCGGGTCTGATGACCACGGCCAAGCACTTCG 4903

Qy     446 TGTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGCATCGTCACGGAGCGGGCTC 505
           |  |||  | |||||| |||  ||| |    ||| |   ||  |||   |||| | | |
Db    4904 CGGCCAACAACCAGGAGAACAACCGCTTCTCCGTGAACGCCAATGTCGACGAGCAGACGC 4963

Qy     506 TCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCCAGCCGGGTGCGT 565
           |||| || ||| |    |   ||||| ||  ||        |  ||    |||   | |
Db    4964 TCCGCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCCAAGGCCGGCGCGGCCTCCT 5020

Qy     566 TCATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGCGAGAACCCTAAATATCTTG 625
           |||||   || ||||| ||| |||| ||   | ||| | |||  |||    |    ||
Db    5021 TCATGTGTGCCTACAACGGCCTCAACGGGAAGCCGTCCTGCGGCAACGACGAGCTCCTCA 5080

Qy     626 ATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCGACTGGTACGGCA 685
           |    |||| || | | | ||||| |   | |||    | |||  |||||||  || ||
Db    5081 ACAACGTGCTGCGCACGCAGTGGGGCTTCCAGGGCTGGGTGATGTCCGACTGGCTCGCCA 5140

Qy     686 CATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGCCCGGACCTCCAC 745
           |  |     ||| || ||| |          ||||||||||  ||||||
Db    5141 C---CCCGGGCACCGACGCCATCACCAAGGGCCTCGACCAGGAGATG------------ 5184

Qy     746 GCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCTTTATCCACGTCA 805
                || |     |||     | ||||| | || ||  ||||||     |  | ||
Db    5185 --------GGCGTCGAGCTCCCCGGCGACGTCCCGAAGGGCGAGCCCTCGCCGCCGGCCA 5236

Qy     806 TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTGCTGCCTCCGGAGTGA 865
           |    | |    ||   |       |  |||| |||  |  | | | |||  | |
Db    5237 AGTTCTTCGGCGA---GGCGCTGAAGACGGCCGTCCTGAACGGCACGGTCCCCGAGGCGG 5293

Qy     866 CGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCCGAAACGGCAGCTCTCCTCCGGA 925
           | | ||   |  | | |  |    ||  |  |      || | |   | ||| ||||
Db    5294 CCGTGACGCGGTCGGCGGAGCGGATCGTCGGCCAGATGGAGAAGTTCGGTCTGCTCCTCG 5353
```

```
Qy    926 AGGTTGGCAACGAGGGCATCGTGCTGCTGAAGAACGAGAACAACGTTCTGCCCTTGAGCA 985
Db   5354 CCACTCCGGCGCCGCGGCCCGAGCGCGACAAGGCGGGTGCCCAGGCGGTGTCCCGCAAGG 5413

Qy    986 AGAAGAAGAAGACGCTGATTGTCGGCCCCAACGCCAAGCAGGCCACATACCACGGCGGAG 1045
Db   5414 TCGCCGAGAACGGCGCGGTGCTCCTGCGCAACGAGGGCCAGGCCCTGCCGCTCGCCGGTG 5473

Qy   1046 GCTCTGCCGCACTCAGGGCCTACTACGCAGTCACTCCCTTTGACGGCCTCAGCAAGCAGC 1105
Db   5474 ACGCCGGCAAGAGCATCGCGGTCATCGGCCCGACGGCCGTCGAC-CCCAAGGTCACCGGC 5532

Qy   1106 TCGAGACGCCGCCATCGTACACCGTCGGCGCCTACACCACCGTTCCTCCCATTCTAGGCG 1165
Db   5533 CTGGGCAGCGCCCACGTCGTCCCGGACTCGGCGGCGGCGCCACTCGACACCATCAAGGCC 5592

Qy   1166 AGCAGTGCCTCACGCCCGACGGCGCTCCGGGCATGCGCTGGAGGGTCTTCAACGAGCCCC 1225
Db   5593 CGCGCGGGTGCGGGTGCGACGGTGACGTACGAGACGGGTGAGGAGACCTTCGGGACGCAG 5652

Qy   1226 CTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCTTCACCAAGACGGACATGCACC 1285
Db   5653 ATCCCGGCGGGGAACCTCAGC-------------------------CCGGCGTTCAACC 5686

Qy   1286 TGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGGTACGCCGACATGGAGGGCACGT 1345
Db   5687 AGGGCCACCAGCTCGAGCCGGGCAAGGCGGGGGCGCTGTACGACGGCACGCTGACCGTGC 5746

Qy   1346 ACACCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGTCGTCTGCGGCACGGCAAAGG 1405
Db   5747 CCGCCGACGGCGAGTACCGCATCGCGGTCCGTGCCACCGGTGGTTACGCCACGGTGCAG- 5805

Qy   1406 CGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCCCCGGCGATGCCT 1465
Db   5806 ------------------------------------------------CTCGGCAGCCACA 5818

Qy   1466 TCTTCGGCTCCGCCACCCGCGAGGAGACGGGCCGCATCAATCTCGTCAAGGGCAACACGT 1525
Db   5819 CCATCGAGGCCGGTCAGGTCTACGGCAAGGTGAGCAGCCCGCTCCTCAAGCTGACCAAGG 5878

Qy   1526 ACAAGTTCAAGATCGAGTTCGGCTCCGCACCCACCTACACCCTCAAGGGCGACACCATCG 1585
Db   5879 GCACGCACAAGCT-----------------------CACGATCTCGGGCTTCGCGATGA 5914

Qy   1586 TCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAAA 1645
Db   5915 GTGCCACCCCGCTCTCCCTGGAGCTGGGCTGGGTGACGCCGGCGGCGGCCGACGCGACGA 5974

Qy   1646 TCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTTA 1705
Db   5975 TCGCGAAGGCCGTGGAGTCGGCGCGGAAGGCCCGTACGGCGGTCGTCTTCGC-------- 6026

Qy   1706 ACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGG 1765
Db   6027 -CTACGACGACGGCACCGAGGGCGTCGACCGTCCGAACCTGTCGCTGCCGGGTACGCAGG 6085

Qy   1766 ACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGG 1825
Db   6086 ACAAGCTGATCTCGGCTGTCGCGGACGCCAACCCGAACACGATCGTGGTCCTCAACACCG 6145

Qy   1826 GCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACG 1885
Db   6146 GTTCGTCGGTGCTGATGCCGTGGCTGTCCAAGACCCGCGCGGTCCTGGACATGTGGTACC 6205

Qy   1886 GCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGG 1945
Db   6206 CGGGCCAGGCGGGCGCCGAGGCCACCGCCGCGCTGCTCTACGGTGACGTCAACCCGAGCG 6265

Qy   1946 GCAAGCTGTCCCTCAGCTTCCCC---------AAGCGCCTGCAGGACAACCCCGCGTTTC 1996
Db   6266 GCAAGCTCACGCAGAGCTTCCCGGCCGCCGAGAACCAGCACGCGGTCGCCGGCGACCCGA 6325

Qy   1997 TCAACTTCCGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTCTACGTCGGGTACA 2056
Db   6326 CAAGCTACCCGGGCGTCGACAACCAGCAGACGTACCGCGAGGGCATCCACGTCGGGTACC 6385

Qy   2057 GGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACA 2116
Db   6386 GCTGGTTCGACAAGGAGAACGTCAAGCCGCTGTTCCCGTTCGGGCACGGCCTGTCGTACA 6445

Qy   2117 CCACTTTTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGAC---GGCAAGCTGAGCGTGT 2173
Db   6446 CCTCGTTCACGCAGAGCGCCCCGACCGTCGTGCGTACGTCCACGGGTGGTCTGAAGGTCA 6505

Qy   2174 CCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCA 2233
Db   6506 CGGTCACGGTCCGCAACAGCGGGAAGCGCGCCGGCCAGGAGGTCGTCCAGGCGTACCTCG 6565

Qy   2234 AGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGG 2293
Db   6566 GTGCCAGCCCGAACGTGACGGCTCCGCAGGCGAAGAAGAAGCTCGTGGGCTACACGAAGG 6625

Qy   2294 TCGAACTGCAGCCCGGCGAGACGAAGGCGGTGAC 2327
Db   6626 TCTCGCTCGCCGCGGGCGAGGCGAAGACGGTGAC 6659
```

RESULT 10
AAZ87319
ID AAZ87319 standard; DNA; 13613 BP.
XX
AC AAZ87319;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae desosamine biosynthetic gene cluster pikB.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX

FH	Key	Location/Qualifiers
FT	CDS	1..809
FT		/*tag= a
FT		/partial
FT		/product= "PikB gene cluster protein #1 (AAY77204)"
FT		/note= "No initiation codon given in the specification"
FT	CDS	806..2014
FT		/*tag= b
FT		/product= "PikB gene cluster protein #2 (AAY80998)"
FT	CDS	2162..6741
FT		/*tag= c
FT		/product= "PikB gene cluster protein #3 (AAY77205)"
FT	CDS	6834..7402
FT		/*tag= d
FT		/product= "PikB gene cluster protein #4 (AAY77206)"
FT		/transl_except= (pos:6837..6841, aa:Gln)
FT	CDS	7492..8205
FT		/*tag= e
FT		/product= "PikB gene cluster protein #5 (AAY77207)"
FT	CDS	complement (7942..8205)
FT		/partial
FT		/*tag= f
FT		/product= "PikB gene cluster protein #6 (AAY77208)"
FT		/note= "No termination codon given in the specification"
FT		/transl_except= (pos:8270..8272, aa:Val)
FT		/transl_except= (pos:8273..8275, aa:Thr)
FT		/transl_except= (pos:8276..8278, aa:Gly)
FT	CDS	complement (10126..11139)
FT		/*tag= g
FT		/product= "PikB gene cluster protein #7 (AAY80999)"
FT	CDS	complement (11271..12149)

```
FT                   /*tag=  h
FT                   /product= "PikB gene cluster protein #8 (AAY77209)"
FT   CDS             complement (12342..13799)
FT                   /*tag=  i
FT                   /product= "PikB gene cluster protein #9 (AAY77210)"
FT   CDS             complement (13706..15043)
FT                   /*tag=  j
FT                   /product= "PikB gene cluster protein #10 (AAY77211)"
FT   CDS             15404..15574
FT                   /*tag=  k
FT                   /product= "PikB gene cluster protein #11 (AAY77212)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77204, AAY77205, AAY77206, AAY77207, AAY77207, AAY77208,
DR          AAY77209,  AAY77210, AAY77211, AAY77212, AAY80998, AAY80999.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 32; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the desosamine
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC   15439, as given in figure 32.
XX
SQ   Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other;

  Query Match              6.7%;  Score 167.2;  DB 21;  Length 13613;
  Best Local Similarity   50.4%;  Pred. No. 9.2e-22;
  Matches  496;  Conservative    0;  Mismatches  473;  Indels   15;  Gaps    3;

Qy   1359 GGACTGCACCTACGAGCTCGGCCTCGTCGTCTGCGGCACGGCAAAGGCGTACGTAGACGA 1418
Db   3516 GGCGTTCAACCAGGGCCACCAGCTCGAGCCGGGCAAGGCGGGGGCGCTGTACGACGGCAC 3575

Qy   1419 CCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCCCCGGCGATGCCTTCTTCGGCTCCGC 1478
Db   3576 GCTGACCGTGCCCGCCGACGGCGAGTACCGCATCGCGGTCCGTGCCACCGGTGGTTACGC 3635

Qy   1479 CACCCGCGAGGAGACGGGCCGCATCAATCTCGTCAAGGGCAACACGTACAAGTTCAAGAT 1538
Db   3636 CACGGTGCAGCTCGGCAGCCACACCATCGAGGCCGGTCAGGTCTACGGCAAGGTGAGCAG 3695

Qy   1539 CGAGTTCGGCTCCGCACCCACCTACACCCTCAAGGGCGACACCATCGTCCCCGGCCACGG 1598
Db   3696 CCCGCTCCTCAAGCTGACCAAGGGCACGCACAAGCTCACGATCTCGGGCTTCGCGATGAG 3755

Qy   1599 CTCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAAATCGAAAAGTCCGT 1658
Db   3756 TGCCACCCCGCTCTCCCTGGAGCTGGGCTGGGTGACGCCGGCGGCGGCCGACGCGACGAT 3815

Qy   1659 CGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTTAACGC---CGACTG 1715
Db   3816 CGCGAAGGCCGTGGAGTCGGCGCGGAAGGCCCGTACGGCGGTCGTCTTCGCCTACGACGA 3875

Qy   1716 GGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGGACCAGCTCAT 1775
Db   3876 CGGCACCGAGGGCGTCGACCGTCCGAACCTGTCGCTGCCGGGTACGCAGGACAAGCTGAT 3935

Qy   1776 TGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGGGCACCCCCGA 1835
Db   3936 CTCGGCTGTCGCGGACGCCAACCCGAACACGATCGTGGTCCTCAACACCGGTTCGTCGGT 3995

Qy   1836 GGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACGGCGGCAACGA 1895
Db   3996 GCTGATGCCGTGGCTGTCCAAGACCCGCGCGGTCCTGGACATGTGGTACCCGGGCCAGGC 4055

Qy   1896 GACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGGGCAAGCTGTC 1955
Db   4056 GGGCGCCGAGGCCACCGCCGCGCTGCTCTACGGTGACGTCAACCCGAGCGGCAAGCTCAC 4115

Qy   1956 CCTCAGCTTCCCC---------AAGCGCCTGCAGGACAACCCCGCGTTTCTCAACTTCCG 2006
Db   4116 GCAGAGCTTCCCGGCCGCCGAGAACCAGCACGCGGTCGCCGGCGACCCGACAAGCTACCC 4175

Qy   2007 CACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTCTACGTCGGGTACAGGTACTACGA 2066
Db   4176 GGGCGTCGACAACCAGCAGACGTACCGCGAGGGCATCCACGTCGGGTACCGCTGGTTCGA 4235

Qy   2067 GTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACACCACTTTTGC 2126
Db   4236 CAAGGAGAACGTCAAGCCGCTGTTCCCGTTCGGGCACGGCCTGTCGTACACCTCGTTCAC 4295

Qy   2127 CTTTTCCAATCTCTCCGTGTCTCACAAGGAC---GGCAAGCTGAGCGTGTCCCTCTCCGT 2183
Db   4296 GCAGAGCGCCCCGACCGTCGTGCGTACGTCCACGGGTGGTCTGAAGGTCACGGTCACGGT 4355

Qy   2184 GAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCAAGCCCCTCCA 2243
Db   4356 CCGCAACAGCGGGAAGCGCGCCGGCCAGGAGGTCGTCCAGGCGTACCTCGGTGCCAGCCC 4415

Qy   2244 AGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTGCA 2303
Db   4416 GAACGTGACGGCTCCGCAGGCGAAGAAGAAGCTCGTGGGCTACACGAAGGTCTCGCTCGC 4475

Qy   2304 GCCCGGCGAGACGAAGGCGGTGAC 2327
Db   4476 CGCGGGCGAGGCGAAGACGGTGAC 4499


RESULT 11
AAT32999
ID   AAT32999 standard; DNA; 2256 BP.
XX
AC   AAT32999;
XX
DT   18-OCT-1996  (first entry)
XX
DE   Chimaeric thermostable beta-glucosidase coding sequence.
XX
```

```
KW   Chimaera; beta-glucosidase; Cellvibrio gilvus; Agrobacterium tumefaciens;
KW   homology; decomposed cellulose; glucose; cellobiose; cellotriose;
KW   cellotetraose; cellopentaose; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..2076
FT                   /*tag=  a
FT                   /note= "sequence derived from Cellvibrio gilvus
FT                          beta-glucosidase gene"
FT   misc_feature    2077..2256
FT                   /*tag=  b
FT                   /note= "sequence derived from Agrobacterium tumefaciens
FT                          beta-glucosidase gene, encodes amino acids
FT                          759-818 of A.tumefaciens beta-glucosidase"
XX
PN   JP08131168-A.
XX
PD   28-MAY-1996.
XX
PF   09-NOV-1994;   94JP-0299049.
XX
PR   09-NOV-1994;   94JP-0299049.
XX
PA   (NORQ ) NORINSUISANSHO SHOKUHIN SOGO.
XX
DR   WPI; 1996-311636/32.
DR   P-PSDB; AAR97199.
XX
PT   Thermostable chimaeric beta-glucosidase - useful for prodn. of
PT   cellobiose from decomposed cellulose
XX
PS   Claim 2; Page 5-8; 9pp; Japanese.
XX
CC   This is the nucleotide sequence encoding a chimaeric beta-glucosidase
CC   enzyme composed of amino acids 1-692 of Cellvibrio gilvus
CC   beta-glucosidase and amino acids 759-818 from the Agrobacterium
CC   tumefaciens beta-glucosidase, replacing amino acids 693-752 of the
CC   Cellvibrio sequence. The chimaera was constructed by isolating the
CC   C.gilvus beta-glucosidase gene and comparing its sequence with the
CC   corresp. genes from several other species, e.g. Hansenula anomala,
CC   Ruminococcus albus, Butyrivibrio fibrisovens, Agrobacterium tumefaciens,
CC   etc. The homologous sequences from the other species could then be used
CC   to generate chimaeric glucosidase genes.  The novel chimaeric protein
CC   acts on decomposed cellulose to generate glucose and cellobiose and has
CC   a reduced ability to decompose cellobiose as compared to cellotriose,
CC   cellotetraose or cellopentaose as substrate.  The novel enzyme has an
CC   optimum pH of 6.0 and temp of 40-45 deg. C, is stable in pH 4-9 and at
CC   35 deg. C.
XX
SQ   Sequence 2256 BP; 370 A; 786 C; 763 G; 337 T; 0 other;

  Query Match             6.4%;  Score 161;  DB 17;  Length 2256;
  Best Local Similarity   52.4%;  Pred. No. 1.1e-20;
  Matches  449;  Conservative    0;  Mismatches  375;  Indels   33;  Gaps    3;

Qy   1598 GCTCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAAATCGAAAAGTCCG 1657
Db   1346 GCGCGATCCAGGCCCAGGCCCCGAATGCGAAAGTCGTGTTCGACGACGGCCGCGACCCAG 1405

Qy   1658 TCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTTAACGCCGACTGGG 1717
Db   1406 CCCGCGCCGCGCGCGTGGCCGCTGGGGCCGACGTCGCCCTGGTCTTCGCCAACCAGTGGA 1465

Qy   1718 AGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGGACCAGCTCATTG 1777
Db   1466 TCGGCGAGGCCAACGACGCCCAGACGCTCGCGCTGCCGGACGGCCAGGAAGAGCTGATCA 1525

Qy   1778 CCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGGGCACCCCCGAGG 1837
Db   1526 CGTCGGTGGCGGGCGCCAACGGGCGCACCGTGGTCGTGCTGCAGACCGGCGGACCGGTCA 1585

Qy   1838 AGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACGGCGGCAACGAGA 1897
Db   1586 CCATGCCCTGGCTGGCGCGCGTTCCGGCCGTGCTGGAAGCCTGGTATCCGGGCACCAGCG 1645

Qy   1898 CGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGGGCAAGCTGTCCC 1957
Db   1646 GCGGCGAGGCGATCGCGAATGTGCTGTTTGGCGCGGTCAATCCGTCCGGCCACCTGCCGG 1705

Qy   1958 TCAGCTTCCCCAAGCGCCTGCAGGACAACCCCGCGTTTCTCAACTTCCGCACCGAGGCCG 2017
Db   1706 CCACCTTCCCGCAATCCGAGCAGCAACTGCCGCGCCCGAAACTCGATGGCGATCCGAAGA 1765

Qy   2018 GGCGCACGCTGTACGGCGAGGACGTCTAC----------------GTCGGGTACAGGTACT 2062
Db   1766 ACCCCGAGCTGCAGTTCGCCGTCGACTACCACGAAGGTGCGGCGGTCGGCTACAAGTGGT 1825

Qy   2063 ACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACACCACTT 2122
Db   1826 TCGACCTGAAGGGCCACAAGCCGCTGTTCCCGTTCGGGCACGGCCTGTCCTACACAACCT 1885

Qy   2123 TTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGACGGCAAGCTGAGCGTGTCCCTCTCCG 2182
Db   1886 TCGCGTACTCCGGTCTGTCCGGCCAGCTCAAGGATGGCCGCCTGCACGTGCGCTTCAAGG 1945

Qy   2183 TGAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCAAGCCCCTCC 2242
Db   1946 TGACCAACACCGGCAACGTGGCCGGCAAGGACGTGCCGCAGGTGTACGCCGCGCCGAT-- 2003

Qy   2243 AAGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTGC 2302
Db   2004 ---GTCCACCAAATGGGAGGCGCCGAAGCGCCTGGCCGCCTGGAGCAAGGTCGCCCTGC 2059

Qy   2303 AGCCCGGCGAGACGAAGGCGGTGACAATCGAGGAGCAGGAGAAGTACGTCGCT------- 2355
Db   2060 TGCCGGGCGAGACCAAGCCGGGCGCGACCGGTACGGCGGTGCTGAAGATCGCTCCTCGCG 2119

Qy   2356 -----GCGTATTTTGATGAGGAGCGGGATCAGTGGTGTGTCGAAAAGGGTGACTATGAGG 2410
Db   2120 ACTTGGCTTACTTCGATGTCGAGGCCGGTCGTTTCCGGGCTGATGCGGGCAAGTACGAGC 2179

Qy   2411 TTATCGTGAGCGACAGC 2427
Db   2180 TGATCGTGGCGGCCAGC 2196


RESULT 12
AAA09469/c
ID   AAA09469 standard; DNA; 50937 BP.
XX
AC   AAA09469;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Streptococcus oleandomycin gene cluster.
XX
KW   oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW   PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW   acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW   macrolactone; antibiotic; motilide; erythromycin; ss.
XX
OS   Streptococcus antibioticus.
XX
FH   Key             Location/Qualifiers
FT   CDS             152..1426
FT                   /*tag=  a
FT                   /label= oleI
FT   CDS             complement (1528..2637)
FT                   /*tag=  b
FT                   /label= oleN2
FT   CDS             complement (2658..4967)
FT                   /*tag=  c
FT                   /label= oleR
FT   CDS             5772..18224
```

```
FT                   /*tag= d
FT                   /label= ORF1
FT                   /product= 8,8a-deoxyoleandolide_synthase_1
FT   CDS             18267..29717
FT                   /*tag= e
FT                   /label= ORF2
FT                   /product= 8,8a-deoxyoleandolide_synthase_2
FT   CDS             29787..40346
FT                   /*tag= f
FT                   /label= ORF3
FT                   /product= 8,8a-deoxyoleandolide_synthase_3
FT   CDS             40625..41830
FT                   /*tag= g
FT                   /label= oleP1
FT   CDS             41878..43158
FT                   /*tag= h
FT                   /label= oleG1
FT   CDS             43163..44443
FT                   /*tag= i
FT                   /label= oleG2
FT   CDS             44433..45173
FT                   /*tag= j
FT                   /label= oleM1
FT   CDS             45251..46411
FT                   /*tag= k
FT                   /label= oleY
FT   CDS             46491..47714
FT                   /*tag= l
FT                   /label= oleP
FT   CDS             complement (47808..49517)
FT                   /*tag= m
FT                   /label= oleB
XX
PN   WO200026349-A2.
XX
PD   11-MAY-2000.
XX
PF   22-OCT-1999;    99WO-US24478.
XX
PR   29-OCT-1998;    98US-0106100.
PR   16-FEB-1999;    99US-0120254.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Betlach MC,  Shah SK,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-365602/31.
DR   P-PSDB; AAY92707, AAY92708, AAY92709.
XX
PT   Recombinant DNA compound encoding oleandolide polyketide synthase for
PT   synthesizing polyketides comprising a coding sequence for a domain of a
PT   loading module or any one of extender modules
XX
PS   Disclosure; Page 14-26; 86pp; English.
XX
CC   This is part of the Streptococcus antibioticus oleandomycin gene cluster.
CC   The oleandolide polyketide synthase (PKS), also known as
CC   8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
CC   (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
CC   "modular" enzyme, where each ORF encodes 2 extender modules and
CC   the first ORF also encodes the loading module. Each module is composed
CC   of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC   acyl carrier protein (ACP) domain. The oleandolide PKS loading module
CC   contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC   glutamine, present instead of the active site cysteine required for
CC   activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC   of polyketide macrolactones through multistep pathways involving
CC   decarboxylative condensations between acylthioesters followed by cycles
CC   of varying beta-carbon processing activities. The macrolide product of
CC   the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
CC   glycosylation to  yield oleandomycin, an antibacterial polyketide. The
CC   invention concerns an isolated recombinant DNA compound, comprising a
CC   coding sequence for a domain of loading module or any one of extender
CC   modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
CC   promoter. Also discussed are recombinant oleandolide PKS in which the
CC   module 1 KS domain is inactivated by deletion or other mutation. In
CC   particular, the inactivation is mediated by a change in the KS domain
CC   that renders it incapable of binding substrate (the KS1-o mutation),
CC   rendered by mutation in the codon for the active site cysteine. The
CC   oleandolide PKS is useful for synthesizing polyketides, which are useful
CC   as antibiotics and motilides. Heterologous expression of oleandolide PKS
CC   in host cells such as Streptomyces coelicolor and S. lividans is also
CC   made possible. Unmodified oleandolide compounds can be provided to
CC   cultures of Saccharopolyspora erythraea and converted to the
CC   corresponding derivatives of erythromycins A-D.
XX
SQ   Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;

  Query Match             5.5%;  Score 138.6;  DB 21;  Length 50937;
  Best Local Similarity   54.0%;  Pred. No. 1.8e-16;
  Matches  354;  Conservative    0;  Mismatches  289;  Indels   12;  Gaps    3;

Qy   1710 CGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGGACCA 1769
Db   3417 CGACGACGGCTCGGAGGACGGGGACCGCACGTCGCTGTCCCTGCCCGGCCGCCAGGACGA 3358

Qy   1770 GCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGGGCAC 1829
Db   3357 CCTGATCAGCGCCGTGGCCGCGGTCAACCCACGGACCGTGGTGGTGCTCAACACGGGCTC 3298

Qy   1830 CCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACGGCGG 1889
Db   3297 CTCCCTCACCATGCCCTGGCTGAGGAAGACCGCCGCGGTGCTCACCATGTGGTACCCGGG 3238

Qy   1890 CAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGGGCAA 1949
Db   3237 CCAGGCCGGCGCCGAGGCGACCACCGCCCTGCTGTTCGGCGACGCCGACCCAGGAGGCCG 3178

Qy   1950 GCTGTCCCTCAGCTTCCC------CAAGCGCCTGCAGGACAACCCCGCGTTTCTCAACTT 2003
Db   3177 GCTGACCCAGACCTTCCCGGCCGACGAGGGGCAGACGCCGTTCGCCGGGGACGCCAGGCG 3118

Qy   2004 CCGCACCGAGGCCGGGCGCACGCTG---TACGGCGAGGACGTCTACGTCGGGTACAGGTA 2060
Db   3117 CTACCCGGGAGTCGATGACCAACTGGACTACTCCGAGGGCATCTACTCCGGCTACCGCTG 3058

Qy   2061 CTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACACCAC 2120
Db   3057 GTACGAACAGCAGGGCGTCCAGCCGCTGTTCTCCTTCGGCCACGGGTTGTGCTACACCTC 2998

Qy   2121 TTTTGCCTTTTCCAATCTCTCCGTGTCTCACAA---GGACGGCAAGCTGAGCGTGTCCCT 2177
Db   2997 CTTCGACTACCGGGACCTGAAGGTGACGGCGACGGCGGACGGCGGGCTCGACGTCTGCTT 2938

Qy   2178 CTCCGTGAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCAAGCC 2237
Db   2937 CACGCTGCGCAACACCGGTACGCGCACCGGCAAGGAGGTCCCGCAGGTCTACGTGGGGCC 2878

Qy   2238 CCTCCAAGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGGTCGA 2297
Db   2877 GTCCCCGCACGTGCGGGTGGCGCAGGCCAAGCGTGCCCTGGCGGCCTACGGGAAGGTCGA 2818

Qy   2298 ACTGCAGCCCGGCGAGACGAAGGCGGTGACAATCGAGGAGCAGGAGAAGTACGTC 2352
Db   2817 GCTGCGGCCGGGCGAGAGCAGGAGGCTGACCCTGCACGTCGAGCGCAGGCCCTC 2763


RESULT 13
AAT93682
ID   AAT93682 standard; DNA; 2166 BP.
XX
AC   AAT93682;
XX
DT   12-MAR-1998  (first entry)
XX
DE   Thermotoga maritima MSB8 glycosidase encoding DNA.
```

```
XX
KW   Glycosidase; thermostable; textile; food processing; pharmaceutical;
KW   detergent; baking; industry; Thermococcus; Staphylothermus;
KW   Pyrococcus; glucose; soluble oligosaccharide; ss.
XX
OS   Thermotoga maritima.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2166
FT                   /*tag=  a
FT                   /product= Glycosidase
XX
PN   WO9725417-A1.
XX
PD   17-JUL-1997.
XX
PF   10-JAN-1997;    97WO-US00092.
XX
PR   13-SEP-1996;    96US-0712612.
PR   11-JAN-1996;    96US-0583787.
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI   Bylina EJ,  Lam DE,  Mathur EJ,  Swanson RV;
XX
DR   WPI; 1997-372858/34.
DR   P-PSDB; AAW34558.
XX
PT   New thermostable glycosidase(s) - from Thermococcus, Staphylothermus
PT   and Pyrococcus, used in the textile, food processing,
PT   pharmaceutical, detergent and baking industries
XX
PS   Claim 4; Fig 5; 82pp; English.
XX
CC   The present sequence encodes glycosidase isolated from Thermotoga
CC   maritima. The enzyme or its encoding nucleic acid sequence is used for
CC   generating glucose from soluble oligosaccharides. The enzyme can be
CC   used in the food processing, pharmaceutical, textile, detergent and
CC   baking industries. The enzyme is also used to treat lactose intolerance,
CC   as a diagnostic reporter molecule, in corn wet milling or in the fruit
CC   juice industry. The enzymes can be used to hydrolyse guar gum to remove
CC   non-reducing terminal mannose residues. The nucleic acids encoding the
CC   enzyme may be used to generate probes to identify similar sequences.
XX
SQ   Sequence 2166 BP; 668 A; 454 C; 578 G; 466 T; 0 other;

  Query Match             5.4%;  Score 135.8;  DB 18;  Length 2166;
  Best Local Similarity   53.5%;  Pred. No. 4.9e-16;
  Matches  284;  Conservative    0;  Mismatches  247;  Indels    0;  Gaps    0;

Qy    207 GCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAGA 266
Db    258 GCTCGCTTCTACCTGGAACAGAGACCTTCTGGAAGAAGTGGGAAAAGCCATGGGAGAAGA 317

Qy    267 GGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCCC 326
Db    318 AGTTAGGGAATACGGTGTCGATGTGCTTCTTGCACCTGCGATGAACATTCACAGAAACCC 377

Qy    327 TCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAGC 386
Db    378 TCTTTGTGGAAGGAATTTCGAGTACTACTCAGAAGATCCTGTCCTTTCCGGTGAAATGGC 437

Qy    387 TGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTTT 446
Db    438 TTCAGCCTTTGTCAAGGGAGTTCAATCTCAAGGGGTGGGAGCCTGCATAAAACACTTTGT 497

Qy    447 GTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGCATCGTCACGGAGCGGGCTCT 506
Db    498 CGCGAACAACCAGGAAACGAACAGGATGGTAGTGGACACGATCGTGTCCGAGCGAGCCCT 557

Qy    507 CCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCCAGCCGGGTGCGTT 566
Db    558 CAGAGAAATATATCTGAAAGGTTTTGAAATTGCTGTCAAGAAAGCAAGACCCTGGACCGT 617

Qy    567 CATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGCGAGAACCCTAAATATCTTGA 626
Db    618 GATGAGCGCTTACAACAAACTGAATGGAAAATACTGTTCACAGAACGAATGGCTTTTGAA 677

Qy    627 TGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCGACTGGTACGGCAC 686
Db    678 GAAGGTTCTCAGGGAAGAATGGGGATTTGGCGGTTTCGTGATGAGCGACTGGTACGCGGG 737

Qy    687 ATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGCCCGG 737
Db    738 AGACAACCCTGTAGAACAGCTCAAGGCCGGAAACGATATGATCATGCCTGG 788

RESULT 14
AAV36911
ID   AAV36911 standard; DNA; 2166 BP.
XX
AC   AAV36911;
XX
DT   21-DEC-1998  (first entry)
XX
DE   Thermotoga maritima MSB8-6G glycosidase gene coding region.
XX
KW   Glycosidase; MSB8-6G; thermostable enzyme; oligosaccharide;
KW   glucose; sugar; baking; textile; detergent; beta-galactosidase; ss.
XX
OS   Thermotoga maritima strain MSB8-6G.
XX
PN   WO9824799-A1.
XX
PD   11-JUN-1998.
XX
PF   08-DEC-1997;    97WO-US22623.
XX
PR   10-OCT-1997;    97US-0949026.
PR   06-DEC-1996;    96US-0056916.
XX
PA   (DIVE-) DIVERSA CORP.
XX
PI   Bylina EJ,  Lam DE,  Mathur EJ,  Swanson RV;
XX
DR   WPI; 1998-362407/31.
DR   P-PSDB; AAW49862.
XX
PT   Glycosidase enzymes from organisms of the genera Staphlothermus,
PT   Pyrococcus and Thermococcus  - for deriving sugar from
PT   oligosaccharides, useful in the e.g. food processing, textile or
PT   baking industries
XX
PS   Claim 1; Fig 5a-b; 92pp; English.
XX
CC   This isolated polynucleotide comprises a coding region for
CC   glycosidase MSB6-6G (see AAW49862) from a Thermotoga maritima MSB8
CC   clone (6G) that grows optimally at 85 degC in high salt medium.
CC   The sequence shows 53% nucleic acid identity to beta-galactosidase
CC   B of Clostridium thermocellum.  The invention provides 18
CC   polynucleotides (see AAV36907-24) coding for thermostable glycosidases
CC   (see AAW49858-75) having glucosidase, alpha-galactosidase,
CC   beta-galactosidase, beta-mannosidase, endoglucanase
CC   or pullulanase activity.  Vectors and host cells are also claimed.
CC   A method is provided for producing the enzymes by recombinant
CC   techniques.  A claimed method for generating glucose from soluble
CC   cell oligosaccharides comprises contacting a sample (selected from
CC   dairy products, fruit juice, detergent, textile, guar gum, animal
CC   feed, plant biomass or waste product) containing oligosaccharides
CC   (selected from maltose, cellobiose, lactose, sucrose, raffinose,
CC   stachyose, verbascose, cellulose, starch, amylose, glycogen,
CC   disaccharides, polysaccharides and pullulan) with one of the
CC   claimed glycosidases such that glucose is produced.
XX
SQ   Sequence 2166 BP; 675 A; 444 C; 581 G; 466 T; 0 other;
```

```
  Query Match             5.4%;  Score 135.8;  DB 19;  Length 2166;
  Best Local Similarity   53.5%;  Pred. No. 4.9e-16;
  Matches  284;  Conservative    0;  Mismatches  247;  Indels    0;  Gaps    0;

Qy     207 GCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAGA 266
           ||||| ||| || |  |||  |   || || ||||| |  || ||    |||||  |||| 
Db     258 GCTCGCTTCTACCTGGAACAGAGACCTTCTGGAAGAAGTGGGAAAAGCCATGGGAGAAGA 317

Qy     267 GGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCCC 326
           |  |  |  .|  |||   ||||| | || |  ||  | || ||||| || |   |||
Db     318 AGTTAGGGAATACGGTGTCGATGTGCTTCTTGCACCTGCGATGAACATTCACAGAAACCC 377

Qy     327 TCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAGC 386
           |||  ||||| |    |||||||        || |||||  ||||  | ||       ||
Db     378 TCTTTGTGGAAGGAATTTCGAGTACTACTCAGAAGATCCTGTCCTTTCCGGTGAAATGGC 437

Qy     387 TGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTTT 446
           | |  ||  |  ||   ||  ||||       || |||   ||    || || |||||| |
Db     438 TTCAGCCTTTGTCAAGGGAGTTCAATCTCAAGGGGTGGGAGCCTGCATAAAACACTTTGT 497

Qy     447 GTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGCATCGTCACGGAGCGGGCTCT 506
           ||  | |||||    |   | ||| | ||| | |  |||||  | ||||| || ||
Db     498 CGCGAACAACCAGGAAACGAACAGGATGGTAGTGGACACGATCGTGTCCGAGCGAGCCCT 557

Qy     507 CCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCCAGCCGGGTGCGTT 566
           | | ||||| ||          ||  | ||||||||     |  |    ||  |  |  |
Db     558 CAGAGAAATATATCTGAAAGGTTTTGAAATTGCTGTCAAGAAAGCAAGACCCTGGACCGT 617

Qy     567 CATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGCGAGAACCCTAAATATCTTGA 626
           ||||  || |||||     | |||||    |  ||     |||||        | |  |
Db     618 GATGAGCGCTTACAACAAACTGAATGGAAAATACTGTTCACAGAACGAATGGCTTTTGAA 677

Qy     627 TGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCGACTGGTACGGCAC 686
           | | ||  |  | |||||||| |  |  ||  |  | ||||||||||||||||
Db     678 GAAGGTTCTCAGGGAAGAATGGGGATTTGGCGGTTTCGTGATGAGCGACTGGTACGCGGG 737

Qy     687 ATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGCCCGG 737
           | |||   |  ||||    |  ||| ||  |||  |    ||||| ||
Db     738 AGACAACCCTGTAGAACAGCTCAAGGCCGGAAACGATATGATCATGCCTGG 788


RESULT 15
AAI99683
ID   AAI99683 standard; DNA; 4403765 BP.
XX
AC   AAI99683;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

  Query Match             4.5%;  Score 111.6;  DB 22;  Length 4403765;
  Best Local Similarity   52.0%;  Pred. No. 2.2e-11;
  Matches  345;  Conservative    0;  Mismatches  289;  Indels   30;  Gaps    3;

Qy    1723 GAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGGACCAGCTCATTGCCGAC 1782
           ||||||  |||| ||||      ||  ||| ||  | |  | |||   ||| || ||||
Db  217690 GAGGGCTTCGACAGCGCCGATCTGTCGCTGCCATGGGGTCAGGATGCGCTGATCGCCGCA 217749

Qy    1783 GTGGCCGCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGGGCACCCCCGAGGAGATG 1842
           || ||  |||| ||| | || ||||| || ||  |  |||||||| ||| | |   |||
Db  217750 GTCGCGTCCGCCAACGCGAATACCGTTGTGGTGCTTGAGACGGGCAACCCGGTGACCATG 217809

Qy    1843 CCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACGGCGGCAACGAGACGGGC 1902
           |||||||  |||| |   | |||| |||| ||||||||||    ||| | | |   |||
Db  217810 CCCTGGCGCGACTCGGTGAACGCCATCATGCAGGCCTGGTATCCGGGCCAGGCGGGTGGC 217869

Qy    1903 AACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGGGCAAGCTGTCCCTCAGC 1962
           |  || |||| ||  | ||    ||  |    || || ||||||  |||| |  ||| |
Db  217870 CAGGCCGTTGCGGAGATTGTGACCGGGCAGGTGAATCCTTCGGGCCGGCTGCCGATCACC 217929

Qy    1963 TTCCCCAAGCGCCTGCAGGACAACCCCGCGTTTCTCAACTTCCGCACCGAGGCCGGGCG- 2021
           |||||        ||      | |  ||     |  |        |     ||  |||  | |
Db  217930 TTCCCGGTCGATCTCGGTCAGACGCCACGCTCGCAACCGCCCGAGCTCGGTGCCCCGTGG 217989

Qy    2022 --------------CACGCTGTACGGCGAGGACGTCTACGTCGGGTACAGGTACTACGAG 2067
                           |  |  |||  ||||| || | | || || ||| | |  |  |
Db  217990 GGGACATCGACCACGATCCACTACACCGAGGGCGCCGATGTTGGTTACCGCTGGTTTGCC 218049

Qy    2068 TTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACACCACTTTTGCC 2127
             |  |  ||  |   | ||| | || || |||||| |||||| |||| ||| |
Db  218050 AGCACAAATCAGACCCCGATGTTCGCGTTCGGTCACGGCTTGTCCTATACCAGTTTCGAG 218109

Qy    2128 TTTTCCAATCTCTCCGTGTCTCACAAGGACGGCAAGCTGAGCGTGTCCCTCTCCGTGAAG 2187
           | |    | ||    ||| |   |  | |  ||   ||  ||      ||  ||| |
Db  218110 TATCGTGACCTGGTGGTGACGGGC---GGCCACACCGTGCACGCCAGTTTCAGCGTTACC 218166

Qy    2188 AACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCAAGCCCCTCCAAGCG 2247
           ||||| |||  |     ||| ||  | ||  | ||||| ||  | |  |    |  | |
Db  218167 AACACGGGCGACCGCAGCGGGGCGGATGTCCCGCAGCTGTATATGATCGCAGCTCCCGGC 218226

Qy    2248 GCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTGCAGCCC 2307
           |    |      ||            | |   ||| ||||    |||||| ||  | |||
Db  218227 GAATCGCGGTTGCG-----------GTTGCTGGGATTCGAGCGGGTCGAGCTCGAACCC 218274

Qy    2308 GGCGAGACGAAGGCGGTGACAATCGAGGAGCAGGAGAAGTACGTCGCTGCGTATTTTGAT 2367
           ||| ||||   |  ||| |  ||||||| | |   |       ||||    ||   |
Db  218275 GGCCAGACTCGGCGGGTAAGGATCGAGGCGGACCCGCGACTGCTCGCCCGCTACGACGGC 218334

Qy    2368 GAGG 2371
           ||||
Db  218335 GAGG 218338
```

Search completed: May 4, 2003, 01:04:22
Job time : 1869.01 secs